Vol. 50 FEBRUARY 1959 No. 1

The Journal of Educational Psychology

CONTENTS

Anxiety, Authoritarianism, and Student Attitude Toward Departmental Control of College Instruction 1
A. W. BENDIG AND PETER T. HOUNTRAS

Speed of Processing Visual Stimuli and Its Relation to Reading 8
LUTHER C. GILBERT

Saccadic Movements as a Factor in Visual Perception in Reading ... 15
LUTHER C. GILBERT

The Relationships Between Various Forms of Aggression and Popularity Among Lower-Class Children 20
GERALD S. LESSER

Some Relations Between Test Scores and Item Statistics 26
FRANCES SWINEFORD

The Comparative Personality Adjustment of Superior and Inferior Readers 31
RALPH D. NORMAN AND MARVIN F. DALEY

Concurrent Validity of the "Warm Teacher Scales" 37
M. STEPHEN SHELDON, JACK M. COALE, AND ROCKNE COPPLE

An Evaluation of the Effectiveness of a Freshman Mathematics Course. 41
J. STANLEY AHMANN AND MARVIN D. GLOCK

AMERICAN PSYCHOLOGICAL ASSOCIATION

The Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

LEE J. CRONBACH
University of Illinois
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
HAROLD E. JONES
University of California
IRVING LORGE
Teachers College, Columbia University
DOROTHEA McCARTHY
Fordham University
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
PERCIVAL M. SYMONDS
Teachers College, Columbia University
GEORGE G. THOMPSON
Syracuse University

The Journal of Educational Psychology publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

SARAH WOMACK
Editorial Assistant

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Mt. Royal & Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D. C.

Second-class postage paid at Baltimore, Md.
Copyright, 1959 by the American Psychological Association, Inc.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 50 February 1959 Number 1

ANXIETY, AUTHORITARIANISM, AND STUDENT ATTITUDE TOWARD DEPARTMENTAL CONTROL OF COLLEGE INSTRUCTION[1]

A. W. BENDIG AND PETER T. HOUNTRAS
University of Pittsburgh

The amount of control exercised by a department over the classroom policies and procedures utilized by college teachers obviously varies considerably from course to course, from department to department, and even among different colleges and universities. The determination of course mechanics, such as textbook selection, reading assignments, testing and grading procedures, and multifarious details that contribute to formal course structure may run the gamut from a *laissez-faire* situation where each teacher is king, selecting the procedures he believes are most appropriate for each of his sections, to a policy of departmentally imposed uniformity in classroom procedures.

Little attention has been paid to the attitudinal response of students to variations in the amount of departmental control over the structure of their courses. McKeachie (1951) and others have argued that the grade-oriented anxiety of college students is more easily controlled and directed toward course achievement in courses that are highly structured so that the student always knows what is expected of him and how the course will be conducted by the instructor. Previous research (Lazarus, Deese, & Osler, 1952; Teevan & McKeachie, 1954) has emphasized the view that the anxiety level of individual *Ss* is negatively related to successful achievement in complex performance situations. This hypothesis of McKeachie suggests that there should be a positive relationship between measures of student anxiety level and favorable attitudes toward the departmental structuring of classroom procedures.

The work of Adorno Frenkel-Brunswik, Levinson, and Sanford (1950) also suggests that the personality trait of authoritarianism may be related to student preferences for departmental control. The highly authoritarian individual dislikes and becomes highly anxious in an unstructured and ambiguous social situation and expresses strong preferences for formal structure. The relationship between test measures of authoritarianism and anxiety has received some attention and research has produced contradictory evidence. Correlations between the F scale devised by Adorno et al. (1950) as a measure of authoritarianism and Taylor's Manifest Anxiety Scale (1953) were reported as positive and high by Jones (1953) and Davids (1955) and as essentially zero by Davids and Eriksen (1957) and Masling (1954). The suggestion by Edwards (1957) that Taylor's MAS and similar scales are heavily saturated with "social desirability" factors in *Ss'* responding to the items further complicates the picture.

The present study was designed to fur-

[1] The authors wish to express their appreciation to Vincent A. Tamburo, undergraduate honors major in psychology, for his valuable assistance in part of the statistical computations.

ther explore the interrelationships among measures of anxiety, authoritarianism, and attitude toward departmental control among undergraduate and graduate students in education.

Procedure

Scales

Three separate psychometric scales, yielding four scores, were assembled into a booklet to be given to each *S* with the *S*'s responses to the items recorded on a single IBM answer sheet. A description of the scales is given below.

Department control. Twenty statements, purporting to measure *Ss*' attitudes toward the amount of control that a college department should exercise over the classroom procedures and policies of the instructor, were written by the authors and revised for clarity and unambiguity. Typical statements were the following:

> The individual instructor should decide when to take attendance in his class rather than the department requiring him to take it once every week.
>
> Departmental policy should determine the kind of tests (essay, true-false, multiple-choice) to be given in a course and not the individual instructor.

The *Ss* were requested to respond to each statement on a three-point scale (mildly agree, neutral, mildly disagree) as to their feelings toward each of the classroom policies as they related to all college courses that the *Ss* had taken or might take in the future, and were instructed not to limit their expressed opinions specifically to the course in which they were currently enrolled. The responses were quantified for scoring by giving two points to the extreme statement (agree or disagree) that, on an a priori basis, represented an opinion favorable toward departmental control over the instructor, while neutral responses were given a weight of one point. Thus a high score on this scale represents an attitude favorable toward departmental control, while a low score presumably reflects a favorable attitude toward more independence for the classroom instructor. The items were approximately counterbalanced for direction of scoring to minimize response bias on the part of the *Ss* with nine statements being scored in the "agree" direction and 11 statements scored in the "disagree" direction. The 20 items were randomized for presentation to the *Ss*. For convenience, we will refer to this scale as the Instructional Control Attitude Scale (ICAS).

Anxiety. The anxiety scores were derived from the administration of Cattell's IPAT Anxiety Scale (1957). The 40 trichotomously-scored items comprising this test are divided into 20 "cryptic" or "subtle" items that have been shown through factor analytic studies to be highly loaded with the "anxiety" factor and 20 "overt symptomatic" or less disguised items that are also heavily loaded with the same factor. Scores on the first set of items give a measure of "covert anxiety," while scores on the second set yield an "overt anxiety" measure. Although both scores measure the same "anxiety" factor, they may have differential validity since the "cryptic" items are presumably less influenced by the *Ss*' tendencies toward giving socially desirable responses to personality items, a factor that has been emphasized by Edwards (1957). High scores on both subscales reflect high degrees of "covert" and "overt" anxiety.

Authoritarianism. 28 items were drawn from Forms 45 and 40 used by Adorno et al., (1950, pp. 255–257) in their studies of the authoritarian personality. Instead of using the cumbersome six-point item response method used by the authors of the statements, a simpler trichotomous response method was used with "agree" responses to each item being given a weight of two points, "undecided" responses a weight of one, and "disagree"

responses a weight of zero. High scores on this F scale presumably measure strong tendencies toward authoritarian attitudes.

The three groups of test items (departmental control, anxiety, and authoritarianism) were assembled into a single test booklet with individual sections labelled "Classroom Policies Questionnaire", "Self Analysis Scale", and "Public Opinion Inventory" being preceded by instructions to the *Ss* on how to record their item responses on the separate IBM answer sheets.

Subjects

The scales were administered to a total of 219 students (104 men and 115 women) enrolled in eight undergraduate and graduate classes on educational psychology. The 109 undergraduate *Ss* (31 men and 78 women) were almost exclusively sophomore pre-education students enrolled in four sections of introductory educational psychology while the 110 graduate *Ss* (73 men and 37 women) were graduate students in the School of Education working toward advanced degrees and enrolled in four sections of courses in human learning and educational research. All eight sections were taught by three male faculty members having a Ph.D. degree in psychology and with several years of college teaching experience. The scales were administered during one class period midway in the semester with the *Ss* requested to indicate only their age, sex, and academic level (undergraduate or graduate) on the answer sheets, thus preserving student anonymity.

Results

To estimate the internal consistency reliability of the 20-item Instructional Control Attitude Scale (ICAS), a random sample of 150 *Ss* was selected from the total group of 219 *Ss* with the only restriction on the randomness of selection being that equal numbers of men and women *Ss* were included in the sample. Of the 150 *Ss*, 79 were graduate students and 71 were undergraduates. The reliability of the ICAS was computed using the variation of Kuder-Richardson Formula 20 suggested by Ferguson (1951) for use with trichotomously-scored items and the resulting reliability coefficient for 150 *Ss* was .66.

Each of the four scale scores obtained were separately subjected to an analysis of variance applied to a 2×2 factorial design with student sex and academic level (undergraduate vs. graduate) as the independent variables. Because of disproportionate subclass frequencies, the method of unweighted means (Snedecor, 1956, pp. 385–386) was used in these analyses. The results are presented in Table 1. The graduate student *Ss* had a significantly (.01 level) lower mean score (10.0) on the ICAS than did the undergraduate *Ss* (13.1), while there was not a significant sex difference on the ICAS. No significant sex or academic level differences were apparent for either of the two anxiety scores (Covert Anxiety and Overt Anxiety). The male *Ss* were significantly (.05 level) more authoritarian (mean = 21.9) than were the female *Ss* (mean = 18.8), and there was a tendency (significant only at the .10 level) for the graduates to be less authoritarian (mean = 19.1) than the undergraduates (mean = 21.5) on the F scale. None of the sex by academic level interaction mean squares approached significance in any of the four analyses.

Product-moment correlations among the four scales were computed within each of the sex groups of the undergraduate-graduate dichotomy. The four interscale correlations (one from each of the sex-level subgroups of *Ss*) from each pair of scales were tested for homogeneity by the chi-square test described by Edwards (1950, p. 135), and, when this test gave no evidence to reject the hypotheses of homogeneous correlations, the coefficients were

TABLE 1

ANALYSES OF VARIANCE OF MEAN DIFFERENCES BETWEEN UNDERGRADUATE AND GRADUATE STUDENTS ON ATTITUDES TOWARD INSTRUCTIONAL CONTROL, ANXIETY, AND AUTHORITARIANISM

Source of Variation	*df*	Instr. Control		Covert Anx.		Overt Anx.		Authoritarianism	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Sex	1	.62	.02	.23	.01	87.50	2.11	452.34	5.21[b]
Academic level	1	432.24	13.01[c]	7.46	.32	36.93	.89	276.41	3.18[a]
$S \times AL$	1	.62	.02	2.26	.10	6.73	.16	15.96	.18
Within subgroups	215	33.23		23.42		41.38		86.80	

[a] $P < .10$
[b] $P < .05$
[c] $P < .01$

TABLE 2

CORRELATIONS OF THE INSTRUCTIONAL CONTROL ATTITUDE SCALE (ICAS) WITH ANXIETY AND AUTHORITARIANISM SCALES

Scales	Undergraduates		Graduates		Homogeneity of Corr. (Chi Square)	Average Correlation
	Male ($N = 31$)	Female ($N = 78$)	Male ($N = 73$)	Female ($N = 37$)		
ICAS vs. Covert Anxiety	−.26	.23[b]	−.15	.08	7.93[b]	
ICAS vs. Overt Anxiety	−.05	.17	−.10	.07	1.09	.03
ICAS vs. Authoritarianism	.27	.26[b]	.41[c]	.09	2.96	.29[c]
Covert Anx. vs. Overt Anx.	.59[c]	.66[c]	.63[c]	.53[c]	.98	.62[c]
Covert Anx. vs. Authoritarianism	.08	.06	.06	.06	.01	.06
Overt Anx. vs. Authoritarianism	.31[a]	.20[a]	.00	.28[a]	3.12	.16[b]

[a] $P < .10$
[b] $P < .05$
[c] $P < .01$

averaged by the usual *r*-to-*z* method (Edwards, 1950, pp. 133–134). The results of these computations can be found in Table 2.

Covert Anxiety and Overt Anxiety showed a consistently high average intercorrelation ($r = .62$) as might be expected since both subscales were designed to measure the same "anxiety" factor. Covert Anxiety is not related to Authoritarianism ($r = .06$) while the average correlation between Overt Anxiety and the F scale ($r = .16$) is significant at the .05 level with three of the four intrasubgroup coefficients being significant at the .10 level.

Scores on the ICAS were significantly (.01 level) related to Authoritarianism for the total group ($r = .29$) with *Ss* high on the F scale preferring more departmental control over course procedures. Although ICAS scores were unrelated to Overt Anxiety ($r = .03$), the pattern of subgroup correlations between the ICAS and Covert Anxiety scales is somewhat puzzling. The chi-square test of the homogeneity of

the subgroup correlations between ICAS and Covert Anxiety was significant at the .05 level (chi-square = 7.93 with three degrees of freedom), precluding the pooling of the subgroups to compute the average correlation between these two scales as was done with other interscale correlations. Critical ratio tests of the significance of the differences between pairs of individual subgroup correlations using the *r*-to-*z* technique described by Edwards (1950, pp. 131–132), indicated that the difference between the interscale correlations for the undergraduate male and female *Ss* (r's = −.26 and .23) was significant at the .05 level ($z = 2.25$), while the difference for the graduate sex subgroups (r's = −.15 and .08), although in the same direction, was not significant ($z = 1.11$). Neither the difference in correlations between the male undergraduate and graduate subgroups (r's = −.26 and −.15), nor the difference between the two female subgroups (r's = .23 and .08) approached statistical significance (z's = .49 and .75). Pooling of the undergraduate and graduate subgroups within each sex group indicated that both average interscale correlations (r's = −.18 and .19) were different from zero, but only at the .10 level of confidence, and that the difference between these two average correlations (Edwards, 1950, p. 136) was significant at the .01 level ($z = 2.64$). Apparently the lack of homogeneity of the four subgroup correlations between the ICAS and Covert Anxiety scales, found in the original chi-square test, reflects a sex difference, and not an academic level difference, that is quite apparent in the undergraduate subgroups and also appears, to a lesser extent, in the graduate *Ss*.

Discussion

The results of the analyses of variance of the four test scores directly contradict the common conception of the graduate student in education as being much more anxious and authoritarian than the undergraduate student. The graduate *Ss* scored significantly (.01 level) lower on the ICAS and there was a tendency, significant only at the .10 level, for them to show a lower mean score on the F scale. In addition, no significant differences between the two academic groups of *Ss* were found on either of Cattell's anxiety scales and both of the F ratios testing the significance of the differences between graduate and undergraduate *Ss* were less than unity. Whether it is the selection process in determining which *Ss* continue on into graduate education classes, the age difference between the two groups of *Ss*, or whether it is the actual experience of teaching that results in lessened authoritarian attitudes in our sample of graduate education students cannot be determined from the data collected for this study.

The hypothesized positive correlation between ICAS and Authoritarianism was confirmed although the relationship was not large ($r = .29$). The only moderate reliability of the ICAS ($r = .66$) probably reduced the apparent extent of the correlation between these two variables. However, this result substantiates the concept that authoritarian verbal responses to general social attitude questions generalize to expressed preferences for departmental control over college classroom procedures. The authoritarian would be happier in a well-structured course where the structure is imposed by the department.

The relationship between attitude toward instructional control and objective measures of anxiety is not as simple as we had originally assumed. The correlation between the ICAS and Overt Anxiety was essentially zero ($r = .03$). The ICAS appeared to be weakly related to Covert Anxiety, but the relationship was reversed between the two sex groups with the correlation being negative for men ($r = .18$) and positive for women ($r = .19$). The

discussion by Edwards (1957) of "social desirability" factors in responding to "subtle" and "obvious" personality test items suggests that the Overt Anxiety scores were contaminated by the "social desirability" variables and this may have obscured the relationship of anxiety, as measured by the Overt Anxiety scale, to instructional control attitudes as measured by the ICAS. The Covert Anxiety scale, presumably being less contaminated by "social desirability," may better reflect the relationship of anxiety and instructional control attitudes. However, the apparent sex difference in the direction of the correlation between the ICAS and Covert Anxiety, if not a result of sampling error, offers a fertile ground for speculation. Covertly anxious men students prefer their instructor to have complete control over course procedures while covertly anxious women students prefer the department to structure the course. This may be due to most college courses having male instructors and male students feeling greater confidence and empathy with their like-sex instructors, whereas female students tend to distrust the control of their predominantly male instructors over course procedures and prefer the institutional restrictions of the neuter department over course structure. Another possibility is that the men *Ss*, being significantly more authoritarian than the women *Ss*, can more easily empathize with and trust the authoritarian figure of the course instructor, regardless of the sex of the instructor, with the more covertly anxious men *Ss* relying more on the more immediate authority figure of the instructor rather than the more nebulous authority control of the department. The covertly anxious women *Ss*, somewhat lower in authoritarianism, may distrust the authoritarian image of the instructor and prefer the socially more distant institutional control of the department. Essentially this hypothesis suggests that authoritarianism level influences the direction of the relationship between covert anxiety and preferences for course control with the correlation between these two variables being negative among authoritarians and positive among nonauthoritarians, regardless of the sex of the *Ss*. Obviously our data do not permit a test of the comparative adequacy of alternative explanations of the significant sex differences in obtained correlations, one based upon a sex difference and the other upon an authoritarianism difference in student perceptions of the role of the college teacher.

Summary

A 20-item scale of student attitude toward departmental control of college teachers (the Instructional Control Attitude Scale) was constructed and found to be moderately reliable ($r = .66$). This scale, along with questionnaires measuring covert anxiety, overt anxiety, and authoritarianism were administered to 219 undergraduate and graduate students in education. The graduate education *Ss* preferred less departmental control of their college teachers and were less authoritarian than were the undergraduate *Ss*. No differences were found on either anxiety scale. Favorable attitudes toward departmental control were (*a*) positively correlated with general authoritarianism ($r = .29$) for all *Ss*, (*b*) negatively correlated with covert anxiety for male *Ss* ($r = -.18$), (*c*) positively correlated with covert anxiety for female *Ss* ($r = .19$), and (*d*) not correlated with the measure of overt anxiety ($r = .03$).

REFERENCES

Adorno, T. W., Frenkel-Brunswik, Else, Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.

Cattell, R. B. *Handbook for the IPAT Anxiety Scale*. Champaign, Ill.: Inst. for Pers. and Ability Testing, 1957.

DAVIDS, A. The influence of ego-involvement on relations between authoritarianism and intolerance of ambiguity. *J. abnorm. soc. Psychol.*, 1955, **51**, 415–420.

DAVIDS, A., & ERIKSEN, C. W. Some social and cultural factors determining relations between authoritarianism and measures of neuroticism. *J. consult. Psychol.*, 1957, **21**, 155–159.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

FERGUSON, G. A. A note on the Kuder-Richardson formula. *Educ. psychol. measmt.*, 1951, **11**, 612–615.

JONES, M. B. Aspects of the autonomous personality: I. Manifest anxiety. *U.S. Naval Sch. Aviat. Med. Res. Rep.*, 1953, Proj. No. NM 001 058.25.03.

LAZARUS, R. L., DEESE, J., & OSLER, SONIA F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

MCKEACHIE, W. J. Anxiety in the college classroom. *J. educ. Res.*, 1951, **45**, 153–160.

MASLING, J. M. How neurotic is the authoritarian? *J. abnorm. soc. Psychol.*, 1954, **49**, 316–318.

SNEDECOR, G. W. *Statistical methods* (5th ed.) Ames, Iowa: Iowa State Coll. Press, 1956.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TEEVAN, R., & MCKEACHIE, W. J. Effects on performance of different instructions in multiple-choice examinations. *Mich. Acad. Sci., Arts, and Letters*, 1954, **39**, 467–475.

Received October 8, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 55, No. 1, 1959

SPEED OF PROCESSING VISUAL STIMULI AND ITS RELATION TO READING

LUTHER C. GILBERT

University of California, Berkeley

Tachistoscopic records have demonstrated repeatedly that both good and poor readers at the college level can identify the visual stimuli of phrases of sense material after exposure for a period which is only a fraction of the time required for the fixation pause in reading easy prose. A possible explanation of the phenomenon appears to lie in the use of the *after image* and the *memory after image* in the tachistoscopic tests. It appears possible that the use of these images tends to give a false measure of the speed and span of visual perception. It also appears possible that maximum use of such images may compensate for certain processing deficiencies.

It has occurred to the writer that the experimental interruption of the use of the retinal images is possible and that this technique can be employed to give a better understanding of the discrepancy between the speed and span of visual perception as measured by the tachistoscopic tests and the speed of reading easy prose.

PURPOSE OF THE INVESTIGATION

The purpose of this study was to investigate the influence of varying the processing times for the first stimulus before the *Ss* were permitted to encounter an interfering stimulus. The data were analyzed (*a*) to determine something of the nature of individual differences in speed of processing visual stimuli and (*b*) to determine relationship between the accuracy of perception and reading ability.

Studies relating to the broad topic of speed and span of visual perception are too numerous to review here in detail. Many of these studies are summarized critically in the reports of Arnoult and Tinker (1939), Huey (1912), Vernon (1931), and Woodworth and Schlosberg (1954).

These earlier studies presented evidence to show that adult *Ss* could see four or five short familiar words in 1/10 or even 1/100 of a second, provided the words formed a simple phrase or sentence and the *Ss* were not required to make a saccadic movement or encounter interfering visual stimuli immediately before or after the flashed material.

In attempting to measure speed of visual perception some investigators had the *Ss* press a key as quickly as the words were recognized. The most commonly used method was to expose the words for a short period and the *Ss* were requested to report the words they recognized after the stimuli were removed from the screen. This method seems to be more a measure of speed of vision than speed of perception, since the procedure permits the *Ss* to continue the processing of the material after the flashed words are removed from the screen. The method fails to identify and evaluate the processing time, and consequently yields inexact measures of speed of visual perception.

In a preliminary study (Gilbert, 1957) using 68 *Ss*, it was found that when two words were flashed on the screen for 1/24 of a second and followed immediately by nonsense letters[1] on the same place on the screen, only 3.68% of the words were correctly reported. Only one *S* reported correctly as many as 20% of the words. Since the writer was interested in measuring the

[1] Letters typed in a random order and in a manner to avoid the formation of sense material.

speed with which the readers could avoid the influence of interfering stimuli on perception of sense material, these nonsense letters seemed to constitute a good test of interference.

Subjects

The *Ss* included 64 college juniors, seniors, and graduates, who were members of classes in educational psychology. Of the 64, 41 were women and 23 men. For all of them, the native language was English.

The rate of reading on a standard level, nonfiction article of approximately 2,300 words ranged from 143 words per minute to 464 words per minute. The average for the group was 262 with a standard deviation of 74.14.

Materials

Reading tests. For measuring the speed and comprehension of reading, a 2,307-word test adapted from one of the University of California weekly broadcasts was used.[2] This test and a second broadcast test,[3] similar in character to the first, were given to 110 *Ss*. The correlation between the two tests was .85 ± .02, which indicates a reliability sufficiently high to justify using it as a group test.

Span of perception tests. The six tests used to measure span of perception were all similar in every way. All the words were short and thoroughly familiar. The phrases gradually increased from one word to five words and from one letter to 25 letter spaces. Each test was composed of five spans, increasing from a one-word span to a five-word span. There were five test items for each span. The material was all typed in lower case on the same IBM executive typewriter using Warren High Gloss paper and black carbon ribbon. The material was then photographed with a Cine-Kodak Special II camera at a distance of nine inches to give a large clear type on the screen. A Bell and Howell projector set at 24 frames per second was used for flashing the material on the screen. Plus-X Reversal Safety film was used in the negative form in order to get black words on soft eggshell-color background. A blank page of the Warren High Gloss paper was photographed using enough film to permit the projector to run for three seconds before the words appeared and two seconds after they had disappeared. This allowed ample time for starting and stopping the projector and at the same time for controlling the illumination on the screen.

The material was projected on a Da-lite screen in a room with the lights turned down to where the *Ss* were just able to see to write their answers, but not dark enough to produce full dark adaptation. Each *S* was asked if he could see clearly the material as it appeared on the screen. The *Ss* who reported that the words blurred were eliminated from the study. Only six *Ss* were lost for this reason. A further check on vision was made by flashing two words on the screen for 2/24 of a second without an interfering stimulus following. The 64 *Ss* reported correctly 98.43 per cent of these words. It seems reasonably certain that defective vision was not an important factor in the findings.

The nonsense letters used as interfering stimuli were typed in lower case on the same typewriter used for typing the words. The letters were typed in a random order and in a manner to avoid the formation of sense material. In each case a sufficient number of nonsense letters was used to extend at least two letter spaces beyond the beginnings and ends of the words.

The three examples of test items which follow show (*a*) the sense material and

[2] Broadcast number 2573, U. E. Number 868, University of California Radio Service, October 29, 1944.

[3] Broadcast Number 1254, U. E. Number 3049, University of California Radio Service, February 24, 1952.

(*b*) the nonsense letters used as interfering stimuli:

Item 1

(a) words
(b) lupytkoie

Item 2

(a) his best shoes
(b) bnvcrmxztghjfedksl

Item 3

(a) they will come home early
(b) zaqwsxcderfvbgtyhjshuiklopzxc

The first test in this group was used to measure the span of perception when the *Ss* were allowed unlimited time for use of the images after the words left the screen. This span is designated the "basic" span. This is the kind of tachistoscopic span commonly reported in the literature and the kind commonly used when training for speed of visual perception. The percentage of words correctly reported on each test as a whole is designated "accuracy" of perception. The other five tests used to measure span and accuracy of perception were all designed in a manner to control the time allowed for processing the words being flashed on the screen. This was accomplished by increasing the time the words were left on the screen before superimposing nonsense letters on the same spot on the screen where the words had been presented. For example, the designation 2–0–2 for the first test means that the words were photographed on two successive frames and were followed without a blank frame by nonsense letters on two successive frames. In other words, the words were left on the screen for 2/24 of a second and were followed immediately by nonsense letters. In each case the nonsense letters were left on the screen for 2/24 of a second. Each control test increased by 1/24 of a second the length of time the material was left on the screen. Thus in the fifth test in the series the words were left on the screen for 6/24 of a second before nonsense letters were superimposed. The exposure time for the nonsense letters was held constant at 2/24 of a second in each test.

Preliminary instructions requested each *S* to fixate at the point on the screen where three small dots appeared and to try to see the words which appeared on the screen. The center dot was placed at exactly the mid-point for the sense material. A number of examples familiarized the *Ss* with the procedure. In each case the dots were left on the screen for exactly the same length of time, i.e. 8/24 of a second and were followed by two blank frames, i.e. 2/24 of a second. This made it possible for the *Ss* to anticipate the words. After each phrase was flashed, time was allowed for the *S* to write the words. The *Ss* were tested in small groups of not more than 25 and no *S* was more than 25 feet from the screen and not more than three chairs from the center of the screen. Further evidence that the words on the screen could be seen was the fact that the group as a whole reported accurately more than 98% of the words up to and including the two-word level on the uncontrolled span.

The same procedure (used for the basic span of perception) was followed with the tests for the controlled spans, with the exception that the instant the sense material left the screen there appeared at the same spot on the screen nonsense letters. The *Ss* were instructed to ignore as far as possible the nonsense letters and to remember the phrases of sense material.

Comparison of the Basic and Control Spans of Perception

Table 1 presents data showing the percentage of the words correctly reported on the basic span and on the various controlled spans for each word level and the test as a whole. It is interesting to note that when the *Ss* used a very narrow span of one and two words nearly all the *Ss* were able to arrive at a very high level

of efficiency even when allowed only 3/24 of a second before the interfering stimulus was superimposed. Also of interest is the fact that, as the span of perception the *Ss* were required to use became greater, the longer the sense material had to be left on the screen before superimposing the nonsense letters in order for the *Ss* to achieve the level of the basic span. The data presented in Table 1 bring into sharp focus the fact that in visual perception such as used in reading sense material, *the fixation pause must be long enough in duration to allow time not only to see but also time to process the visual stimuli.* It may well be that the individual differences in speed of processing visual materials is an influential factor in both the span of perception and the length of the fixation pauses used in reading easy prose. If the data in Table 1 were presented graphically they would show the group curves of the controlled spans crossing the lines for the basic spans at various time levels. The group curve for the controlled test of accuracy of perception crosses the line for the basic test between the levels of 5/24 and 6/24 of a second. It is interesting to note that this point of crossing is close to the average length of the fixation pauses commonly found for mature readers reading simple prose.

The data in Table 1 show a marked increase in the percentage of words reported from the 2/24 to 6/24 of a second, with the greatest gain taking place between the 2/24 and the 3/24 time units. At the level of 4/24 of a second the group had reached a level of efficiency on the control span which was slightly above the basic spans of the three words or less. By the level of 5/24 of a second the control span of four words had equaled or exceeded the four-word basic span. By the level of 6/24 of a second the imposed stimuli had no measured detrimental influence on the groups's five word span of perception. However, some of the very slow readers were still experiencing some interference.

These data also indicate that the non-interference times of 4/24 or 5/24 of a second following the flashing of the words on the screen constitute a very important factor in the span of perception. Furthermore they indicate that the time words are left on the screen is not the only factor which should concern us in studying speed and span of perception. Note that words left on the screen for 2/24 of a second and not followed by other visual

TABLE 1

MEAN PERCENTAGE OF WORDS CORRECTLY REPORTED ON THE VARIOUS SPANS (64 *Ss*)

Length of Span in Words	Time Units—Percentage Correct in Each Unit					
	2-0-2	3-0-2	4-0-2	5-0-2	6-0-2	2-0-0 (Basic Span)
One Word	81.25	97.81	98.75	100.00	99.68	99.37
Two Words	86.71	97.81	99.06	99.53	99.68	98.43
Three Words	64.44	93.74	96.24	98.22	96.56	95.94
Four Words	56.14	83.82	85.54	93.04	92.65	91.63
Five Words	36.06	64.06	66.50	71.93	81.32	75.56
Percentage of total	58.38	82.01	84.01	88.36	91.14	88.56
SD	17.76	12.94	10.54	9.50	9.64	10.82

stimuli were as accurately reported as words left on the screen for 5/24 of a second and followed by interfering visual stimuli. Both the length of time the words are left on the screen *and* the length of the period free from interfering stimuli are important factors. These data suggest the possibility that some readers may use part of their fixation time to avoid interference from a new stimulus during the period they need free for processing the visual stimulus. In other words, part of the fixation time may be preventive in nature.

Comparison of Rate of Reading and Speed and Accuracy of Perception

For the purpose of showing the differences in rate of processing the sense material for readers of different levels of ability the group of 64 *Ss* was ranked from the best to the poorest on the gross reading rate and then divided into fourths. The best 25% are designated as S_1 and the poorest 25% are designated as S_4. (S_2 and S_3 constitute the middle 50%.) Table 2 shows the difference between the slow reading group (S_4) and the fast readers (S_1) in speed of processing sense materials. When the sense material was left on the screen for 3/24 of a second, then followed by an interfering stimulus, the S_1 group got a considerably higher percentage of the material than the S_4 group did when the sense material was left on the screen for 6/24 of a second, then followed by an interfering stimulus. Another point of interest in these data is the fact that the S_4 subjects experienced greater difficulty in avoiding interference from the interrupting stimuli than was true for the S_1 students.

Correlation Between Speed and Accuracy of Perception and Rate of Reading

This positive association between rate of processing simple phrase material and the gross reading rate of simple prose material is brought out a little more clearly in Table 3 which shows the coefficients of correlation for the percentage of words processed on each test of perception and the gross reading rate. The magnitude of these correlations with the gross reading rate decreases in an irregular pattern from the 2–0–2 time unit to the 6–0–2 time unit level. It should be remembered that the average accuracy for the 6–0–2 time unit test is similar in magnitude to

TABLE 2

MEAN PERCENTAGE OF WORDS CORRECTLY REPORTED IN THE VARIOUS TIME UNITS BY S_1 AND S_4 GROUPS[a]

(64 *Ss*)

Subjects	Time Units					Basic Span
	2-0-2	3-0-2	4-0-2	5-0-2	6-0-2	2-0-0
S_1 Best 25% of the Readers	78.49	93.07	92.41	96.08	97.17	94.56
SD	14.14	4.66	5.07	3.14	3.02	5.53
S_4 Poorest 25%	39.74	71.74	77.60	81.20	86.91	83.41
SD	19.15	13.28	11.81	8.12	6.96	8.50
Mean Difference	38.75	21.34	14.75	14.84	10.26	11.15
t[b]	6.4	5.9	4.5	5.8	5.2	4.3

[a] Ranked on gross reading rate.

[b] All the *t* ratios are significant at better than the .01 level.

that on the basic span test and at this time level there seemed to be very little influence resulting from the nonsense letters as reflected in the group average. (However, a few of the poor readers were still experiencing some interference.) Therefore it is not surprising that this correlation between scores on the 6–0–2 test and gross reading rate is of a similar magnitude as that found between the basic span test and gross reading rate.

Table 3 also reveals that the magnitude of the correlations between the speed and accuracy of processing the phrases and the effective reading rate[4] are slightly larger but similar in magnitude to those reported for the gross reading rate. The effective reading rate seems to reflect a little better than the gross reading rate the speed with which the *Ss* can process the stimuli which they receive from each fixation pause. If this assumption is true, it would seem logical that the correlations between effective reading rate and speed and accuracy of processing the phrases would be a little higher than for the gross rate and speed and accuracy of processing the stimuli. However, since the correlation between the gross and effective rates for this group of *Ss* was .86 ± .012, no great differences in the magnitude of the correlations between gross rate and processing of phrases and effective rate and processing of phrases should be anticipated.

TABLE 3

RELATION OF PERCENTAGE OF WORDS CORRECTLY PERCEIVED TO GROSS READING RATE AND EFFECTIVE READING RATE

(64 *Ss*)

	Time Units					Basic Span
	2-0-2	3-0-2	4-0-2	5-0-2	6-0-2	2-0-0
	Gross Reading Rate					
r	.63	.58	.48	.56	.22	.32
SE	.076	.083	.097	.087	.019	.080
	Effective Reading Rate					
r	.68	.63	.57	.58	.30	.23
SE	.067	.075	.085	.084	.114	.120

Eye Movements

The evidence presented in the literature on visual perception in reading establishes quite clearly the tendency for slow readers, in reading simple prose, to use a smaller span of visual perception and a longer fixation pause than do faster readers. There are, however, many exceptions to this tendency. These exceptions are doubtless, in many instances, due to the interaction of certain of the major factors involved. For example, the evidence on speed and accuracy of processing visual stimuli of simple phrases of sense material shows that short phrases can be processed faster and with fewer errors than longer phrases. Consequently certain slow readers may, by keeping their units of visual perception very small, be able to read simple prose with shorter than average fixation pauses. The record of Subject R is a good illustration of this type of reading pattern. She read simple prose at the rate of 186 words per minute, with good comprehension. Her eye movement records show that she used 80 fixations per 100 words with average fixation pauses of 6.4 thirtieths of a second. Her fixation pauses were much shorter than the average for slow readers. On the speed of perception tests she demonstrated a high degree of speed and accuracy of processing short phrases of sense material, but was very slow and inaccurate in processing long phrases. The record of Subject P reveals a very different pattern. This *S* read simple prose at 176 words per minute. Her eye movement records show that she used

[4] The effective reading rate is computed by multiplying the gross number of words read by the percentage of correct answers on the test of comprehension.

53 fixations per 100 words, but with an average fixation pause of 9.4 thirtieths of a second. Her wide span of perception and unusually long fixation pauses reveal a very different pattern from that of R. The speed of perception tests reveal that P needs a considerably longer interval of time free from interfering stimuli than the average college student in order to make use of her larger than average span of visual perception.

The fixation pauses in reading serve three purposes. First, the eyes are much more efficient in transmitting the visual stimuli to the cortex when at rest than when in motion. Therefore, the eyes are stopped along the line of print in order to achieve maximum functional efficiency. Second, in order to achieve maximum efficiency in processing the visual stimuli, the retina or cortex needs an interval of time free from interfering visual stimuli. The length of this uninterrupted period is determined by the length of the fixation pause. Individual differences in the length of time needed for retinal or central processing of the visual stimuli no doubt account in part for individual differences in the length of the fixation pauses in reading simple prose. The first two purposes are prerequisites to the third, namely, providing time needed to comprehend the ideas and relationships involved. The substantial correlation .50 ± .14 between the speed and accuracy of visual perception as measured by Test 2–0–2 and the length of the fixation pauses in reading simple prose seems to support this theory.

Summary of Findings

The findings of the investigation may be summarized as follows:

1. Among college students the span of visual perception is unequally influenced by an increased restriction in the period of freedom from interfering visual stimuli following the presentation of simple phrase material.

2. For the average college student, if the phrases are left on the screen for 1/5 or 1/4 of a second before the extraneous visual material is presented, the extraneous visual material has little influence on the span of visual perception.

3. In exceptional cases, *Ss* needed more than 1/4 of a second to be able to avoid the interference of the extraneous material.

4. The narrower the span of perception the easier it is to avoid the influence of the extraneous visual material.

5. Interfering stimuli have a greater influence on the span of visual perception for the slow readers than they do for the fast readers.

6. The greater the degree of freedom from interfering stimuli the lower the correlation between span of visual perception and rate of reading.

7. There is a substantial correlation between the length of the fixation pauses *Ss* use in reading simple prose material and the speed with which the *Ss* can process tachistoscopically-presented stimuli resulting from simple phrases.

REFERENCES

ARNOULT, D. C., & TINKER, M. A. The fixational pause of the eyes. *J. exp. Psychol.*, 1939, **25**, 271–280.

GILBERT, L. C. Influence of interfering stimuli on perception of meaningful material. *Calif. J. educ. Res.*, 1959, **10**, 15–23.

HUEY, E. B. *The psychology and pedagogy of reading*. Macmillan, 1912.

VERNON, M. D. *The experimental study of reading*. Cambridge Univer. Press, 1931.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. Henry Holt, 1954.

Received May 30, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 1, 1959

SACCADIC MOVEMENTS AS A FACTOR IN VISUAL PERCEPTION IN READING

LUTHER C. GILBERT
University of California, Berkeley

It is the purpose of this investigation to study the saccadic movements of the eyes in their relationship to visual perception. Other data in the literature of the subject suggest the theory that effectiveness in reading is determined largely by the functional efficiency of the central nervous system, but there remains the question, "Does the level of functional efficiency of the eyes condition the activity of the central nervous system?"

More specifically it is the purpose of the present study to explore the relationship between the speed and accuracy of perception among college *Ss* in reading simple prose both with and without saccadic movements of the eyes. There is the possibility that the functional efficiency of the eyes not only reflects the activity of the central nervous system but in a way conditions it as well. It is the purpose, therefore, to put this theory to the test.

OTHER STUDIES

In 1930, Vernon published the report of a study in which she attempted to gauge the efficiency of ocular muscles in the execution of steady fixations and lateral movements and to relate this efficiency to the reading process. Her *Ss* were nine adults, eight of them university graduates. She concluded from her data that individual types of motor processes in reading are originally based upon underlying motor ability and upon the variation of the individual movements which make up the habit according to the ability of the eye muscles. She further stated that the use of frequent short pauses, or of less frequent long pauses, and the tendency to overrun the word and then regress to it, are permanent oculomotor habits, unconnected with perception and assimilation of the reading content.

Tinker, in a study reported in 1938, photographed the eye movements of 64 university sophomores in an attempt (*a*) to measure the accuracy of oculomotor control as manifested by the accuracy of voluntary fixation during successive perceptual acts, (*b*) to measure the speed of convergence and divergence during successive visual fixations, and (*c*) to determine the relation of these measures to speed of reading and eye movement measures of reading. He administered the Minnesota Speed of Reading Test and photographed the eye movements during the reading of easy prose. In addition, he photographed successive fixations which the *Ss* made while reading numbers located at the ends of blank horizontal lines 4½ inches long and 6 inches long.

Tinker concluded that for these *Ss* as a group, neither convergence, divergence, nor oculomotor efficiency in fixating numbers at the end of a line-length sweep of the eyes bore any significant relation to proficiency in reading, but that when extreme cases in the group were compared there appeared to be a slight, consistent relation between motor efficiency of the eyes and reading ability. However, this study is limited by the fact that oculomotor control was measured by having the *Ss* fixate numbers at the ends of long lines. The only similar saccadic movement in normal reading is the long sweep of the eyes from the end of one line to the beginning of the next line. These fixations represent a small fraction of the total number of fixations used in reading.

Some years ago the writer of this article studied the oculomotor growth of college

students and of *Ss* from Grades I through IX (Gilbert, 1953). Oculomotor ability was measured by eye movement records of the *Ss* while they tried to fixate single digits distributed at unequal distances along four-inch blank horizontal lines. The evidence from these records suggests that there is a growth in this type of oculomotor ability, with the greatest amount taking place in the primary grades. Efficiency in fixating these single digits is positively associated with the rate of reading simple prose.

The underlying causal factor, or factors, in the positive association of fixation pauses in visually fixating these single numerals and fixation pauses used in reading easy prose has not been clearly identified. Some neurologists (Nielsen, 1951) hold to the theory that the mental processing of the visual stimulus of the number *9* involves a different part of the brain from that used in processing the visual stimulus of the word *nine*. Until this theory is carefully evaluated, caution should be exercised in interpreting these as identical mental functions. Certainly, the fixating of a single numeral (following a saccadic movement) as a method of studying oculomotor efficiency gives an indirect and uncertain measure of the influence of saccadic movements in reading easy prose.

In a more recent study (Gilbert, 1959) the writer presented evidence pointing up the wide range of individual differences among college students in the speed with which they process the visual stimuli of phrases and sentences of sense material when the eyes are functioning without the necessity of making saccadic movements immediately preceding or following the flashing of the sense material. A positive correlation between the speed and accuracy of processing the sense material and the rate of reading easy prose material supports the generally accepted theory that the effectiveness of the processing skills exercises a profound influence upon reading. However, there remains the question, "Is this the whole story? Is it possible that the saccadic movements may exercise an uneven influence on the speed of reading easy prose?" This question is not answered in the present body of literature.

The foregoing studies in oculomotor efficiency differ from the present one in two important respects: (*a*) the material and procedure used, and (*b*) the techniques and analysis.

Subjects

The *Ss* included 76 college juniors, seniors, and graduates, who were members of classes in educational psychology. Of the 76 *Ss*, 29 were men and 47 women. For all of them, the native language was English.

The rate of reading on a standard level, nonfiction and nontechnical article of approximately 2,300 words ranged from 126 words per minute to 602 words per minute. The average for the group was 263 words per minute with a standard deviation of 70.80.

The vision of the group was tested by asking each *S* if he or she could see clearly the words as they appeared on the screen. Any who reported that the words blurred were eliminated from the study. Only 8 *Ss* were lost for this reason. A further check on the vision was made by flashing two-word phrases (all simple, short words) on the screen for 1/24 of a second without a competing stimulus following. The 76 *Ss* reported correctly 97.5% of the words. It seems reasonably certain that neither defective vision nor lack of familiarity with the simple words used in the sentences was an important factor in the results.

Materials

Two tests (I and III) were designed to measure speed and accuracy of visual perception while reading simple prose sentences without the necessity of moving the eyes. Test I was composed of 20 sentences,

each eight words in length—a total of 160 words. In order to avoid as far as possible the influence of unfamiliarity with the vocabulary and sentence structure, the material was written in simple sentences, using only short and thoroughly familiar words. Each sentence was an independent unit of thought. The material was typed on an IBM typewriter using Warren high gloss paper. Each sentence was broken into four phrases of two words each. In order to facilitate photographing of the material, each pair of words was typed on a separate page.

To prevent the necessity of moving the eyes, *three small dots were photographed on eight frames of 16 mm. film.* These dots indicated to the *S* where the words would appear. Two blank frames of the same illumination followed the dots to signal the appearance of the words. Each pair of words appeared at the designated spot on the screen and remained on the screen for 4/24 of a second. Each *unit* was separated from the others by 2/24 of a second. At the completion of each sentence, time was allowed for the *S*s to write down what they saw. The score was one point for each word correctly reported. A perfect score was 160 points.

Example of Test I

The three dots (. . .) were projected on the center of the screen to indicate where each of the four units, or word pairs would be presented. For example, the word pairs, *two boys, will visit, our home, next week,* were all projected to the same position on the screen. Each pair of words was exposed for 4/24 of a second and followed by 2/24 of a second of blank screen.

Example of Test II

With one major exception Test II was similar to Test I. In Test II the words were presented along what might be thought of as a line of print. In order to read the sentence the *S* had to cover what was equivalent to a nine-word line. The time intervals were the same. Each pair of words was exposed for 4/24 of a second and followed by 2/24 of a second of blank screen. The following is an example of the items in this test:

"Our school•will have•open house•this year"
(•separated by three letter spaces)

In Tests II and IV, the dots were used to indicate when and where the first pair of words would appear. Since the positions on the screen where the phrases were presented were fixed, no dots were used for the other phrases.

Test III was similar to Test I except for the fact that the words were left on the screen for 2/24 of a second. Test IV followed the design of Test II, except for the fact that the words were exposed for 2/24 of a second.

Findings of the Investigation

Data concerning the influence of saccadic movements of the eyes on speed and accuracy of visual perception in reading simple prose sentences for the experimental group of 76 cases are presented in the first column of Table 1. These data show that when the sentences were presented two words at a time in a manner which required saccadic movements and each pair of words was exposed for 2/24 of a second the group reported accurately 64% of the material. When comparable material was read at the same speed without eye movements the group averaged 92% correct. Reading at this speed, the difference between the mean percentage of accuracy with and without movements was nearly 28. When the material was left on the screen for twice as long, (4/24) the difference between the mean percentages was about 12. The t ratio between the loss at the 2/24 level and the loss at the 4/24 level is 8.88, which is significant at better than .001 level. It seems that the closer the length of exposure time for the word pairs

TABLE 1

INFLUENCE OF SACCADIC MOVEMENTS OF THE EYES ON SPEED AND ACCURACY OF VISUAL PERCEPTION IN READING SIMPLE PROSE SENTENCES

	Time Units 2–2–2			Time Units 4–2–4		
	Total group 76 *Ss*	S_1[a] group 19 *Ss*	S_4[b] group 19 *Ss*	Total group 76 *Ss*	S_1 group 19 *Ss*	S_4 group 19 *Ss*
Mean Percentage of Words Correctly Reported—No Saccadic Movement	92.03	97.93	84.61	96.32	99.28	93.16
SD	10.55	3.00	12.97	5.10	1.57	6.39
Mean Percentage of Words Correctly Reported—Saccadic Movement	64.35	79.80	52.57	84.56	93.98	75.14
SD	20.82	16.61	23.49	16.39	7.83	21.78
Difference Between Means—Saccadic Movement and no Saccadic Movement	27.68	18.13	32.04	11.76	5.30	18.02
t[c]	10.27	4.57	6.32	5.94	2.82	3.37

[a] Best 25% of the readers.
[b] Poorest 25% of the readers.
[c] All the t ratios are significant at better than the 1% level.

approaches the average duration of the fixation pauses used in reading simple prose the less is the loss in visual perception which is associated with saccadic movements.

For the group as a whole the number of errors in visual perception associated with saccadic movements is decreased as the length of the exposure is increased. Is this equally true for both good and poor readers? In order to check this point the 76 cases were ranked on gross reading rate from the fastest reader to the slowest reader. The best 25% of the readers was compared with the poorest 25%. The best group was designated as S_1 and the group of the poorest readers was designated as S_4. The data in Table 1 show a number of interesting things. Obviously, both good and poor readers make more perceptual errors at both levels of speed when they read with eye movements than they do when reading without movements. The saccadic movements are associated with a substantially greater loss in visual perception for the poor readers at both the 2/24 and 4/24 levels than they are for good readers. The t ratios are 2.784 for the 2/24 level and 2.917 for the 4/24 level. Both are significant at better than the .01 level. Of special interest is the fact that both good and poor readers make fewer perceptual errors reading at the 2/24 level without saccadic movements than they do reading comparable material at the 4/24 level with saccadic movements. Another point of special interest is the evidence that both S_1 and S_4 readers can process this material mentally at a faster rate and more accurately than they actually do when reading with saccadic movements.

From these data one cannot be certain just why visual perception is so much better when the eyes function without saccadic movements than it is following saccadic movements. Neither is it clear why some *Ss* suffer a much greater loss in visual perception following saccadic move-

ments than is true for other *Ss*. Contrary to traditional theory, it appears quite possible that a very real part of this phenomenon is due to individual differences in functional motor efficiency of the eyes. Part may be due to individual differences in speed and accuracy for processing visual stimuli. There remains also the possibility that these factors are interacting.

Summary of Findings

The findings of the investigation may be summarized as follows:

1. These data indicate that both good and poor readers make more perceptual errors when they read with eye movements than they do when reading without movements.

2. Both good and poor readers make fewer perceptual errors reading at the 2/24 second level without saccadic movements than they do reading at the 4/24-second level with saccadic movements.

3. Saccadic movements are associated with a substantially greater loss in visual perception for the poor readers than they are for the good readers.

4. Slow readers seem to need, as a group, a longer interval of time to stabilize their fixations to a point of maximum efficiency than is necessary for the fast readers.

5. Both good and poor readers can process simple prose material mentally at a faster rate and more accurately than they actually do when reading with saccadic movements.

REFERENCES

Gilbert, L. C. Functional motor efficiency of the eyes and its relation to reading. *Univer. Calif. Publ. Educ.*, 1953, **11**(3), 159–232.

Gilbert, L. C. Speed of processing visual stimuli and its relation to reading. *J. educ. Psychol.*, 1959, **50**, 8–14.

Nielsen, J. M. *Clinical neurology*, Paul B. Hoeber, 1951.

Tinker, M. A. Motor efficiency of the eye as a factor in reading. *J. educ. Psychol.*, 1938, **29**, 167–174.

Vernon, M. D. *The movements of the eyes in reading*. London: Med. Res. Council, Spec. Ser. No. 48, 1930.

Received May 30, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 1, 1959

THE RELATIONSHIPS BETWEEN VARIOUS FORMS OF AGGRESSION AND POPULARITY AMONG LOWER-CLASS CHILDREN

GERALD S. LESSER

Educational Clinic, Hunter College

In the study of socialization, there has been considerable emphasis upon the reactions of parents (Davis, 1944; Sears, Whiting, Nowlis, & Sears, 1953; Whiting & Child, 1953) and teachers (Appel, 1942; Jersild & Markey, 1935; Levin, 1955) to the aggressive behavior of children. However, only a limited amount of descriptive and normative information (Cunningham et al., 1951; MacRae, 1954; Piaget, 1932; Tuddenham, 1951) exists concerning the exact nature of the reactions of the child's peers to his aggressive behavior. In general, there has been little empirical specification of the social behaviors which are rewarded and punished by the peer group.

The nature of the peer group's influence as socializing agents depends upon a large variety of variables. The application of rewards and punishments by the peer group differ greatly for different drive areas (e.g., sex, achievement, dependency, aggression, etc.). Also, the use of rewards and punishments by peers in any single drive area, such as aggression, undoubtedly varies with the age, sex, and social class position of the children, among other variables. Social class influence upon reactions to aggression has been especially emphasized. Davis' (1944) report that aggression among lower-class children is learned as a socially rewarded form of behavior has received considerable attention.

However, the conclusion that an over-all positive relationship exists in lower-class groups between aggression and cultural approval appears to be an over-simplification. Since aggression is not a simple, unidimensional variable, it is plausible to assume that the acceptability of the aggressive responses to a lower-class peer group will depend upon the precise form of the aggression manifested. Which aggressive responses are accepted and positively reinforced by peers in lower-class groups and which aggressive responses are rejected and negatively reinforced by peers in lower-class groups? This question indicates the focus of the present study.

There is clear evidence (Cunningham et al., 1951; Murphy, 1937) that groups of young children are capable of making very fine discriminations among the specific responses (even within a single drive area) of their age-mates; these responses are then differentially rewarded or punished by the peer group's behavior. Five different manifestations of aggressive behavior were defined in this study which, on a priori grounds, appeared to be conceptually independent. The hypothesis proposed here is that these different manifestations of aggression meet with different degrees of approval and disapproval by the lower-class peer group.

METHOD

The *Ss* were 74 white boys (ages 10–0 to 13–4) drawn from three fifth grade and two sixth grade classes in three public schools in New Haven, Connecticut.[1] The Kuhlmann-Anderson intelligence quotients of these boys ranged from 78 to 125, with a mean of 103. The three schools are in adjacent districts and the families

[1] The writer wishes to thank Esther Viets, J. Allen Hickerson, May White, John Wesolowski, Margaret Fitzsimons, and the teachers of Barnard, Scranton, and Roger Sherman schools in New Haven, Conn. for their helpful cooperation.

of the children constitute a relatively homogeneous upper lower-class group, according to the Index of Status Characteristics criteria of occupation, source of income, house type, and dwelling area (Warner, Meeker, & Eells, 1949, pp. 121–230). The town of New Haven had been previously mapped for social class location;[2] this mapping confirmed the social class placement of the families of the children in this study.

The variable of popularity was measured through a standard sociometric measure. The children were asked to "List the 3 boys in your class that you would like to have for your best friends." and to "List the 3 boys in your class that you wish were not in your class at all." A popularity score was obtained for each *S* by subtracting the number of times he was mentioned for the second ("enemies") question from the number of times he was mentioned for the first ("friends") question. The Spearman rank-difference correlation coefficients (rho) between the boys' popularity choices and girls' popularity choices ranged from +.74 to +.24 for the different classes; the boys' and girls' popularity choices were therefore combined to provide a single measure of popularity for each *S*.

The measures of overt expression of aggression were obtained through a modified sociometric device, the "Guess Who" technique. Each child in the class received a booklet containing a series of written descriptions of children and was instructed to identify each of these descriptive characterizations by naming one or more classmates. Both the boys and girls in each class completed this booklet; however, all *Ss* were instructed to list only the names of boys in the class in answering the items.

The following instructions appeared upon the cover of the booklet and were also read to the children by the experimenter:

> In this booklet are some word pictures of boys in your class. Read each one and write down the names of the boys whom you think the picture fits.
>
> *Remember:*
>
> Any one picture may fit more than one boy. You may write as many names as you think belong under each picture.
>
> The same boy can be mentioned for more than one picture.
>
> You will have as much time as you need to finish. Do not hurry.

Before beginning the task, the children were also told that only the experimenter (and not the teacher) would see their entries in the booklets and that the task had nothing whatever to do with their school grades. The children were not asked to put their names on the booklets until the end of the session.

The definitions of the five categories of aggressive activity, and the sociometric items used in each category are as follows:

1. Provoked Physical Aggression:
   Definition: to physically attack or injure after provocation.
   Items:
   a. Here is a boy who will fight, but only if someone picks on him first.
   b. Here is someone who will always fight back if you hit him first.
2. Outburst Aggression:
   Definition: to display uncontrolled, "temper tantrum" aggressive behavior.
   Items:
   a. Here is a boy who gets so mad at times that he doesn't know what he is doing.
   b. Here is someone who flies off the handle right away and is very hotheaded.
   c. This boy gets very, very mad at times.
3. Unprovoked Physical Aggression:
   Definition: to physically attack or injure without provocation.

[2] The map of residential areas of New Haven, Connecticut, 1950, was compiled by Jerome K. Myers, under the auspices of USPHS grant MH 263, principal investigators A. B. Hollingshead and F. C. Redlich, Yale University. Copyrighted by Jerome K. Myers.

Items:
a. This boy starts a fight over nothing.
b. Here is someone who is always looking for a fight.
c. Here is a boy who gets mad while he is playing and ends up in a fight.

4. Verbal Aggression:
Definition: to verbally attack or injure.
Items:
a. This boy often threatens other boys.
b. This boy is always scolding when he is playing a game with other boys.
c. This boy makes up stories and lies about other children in the class.

5. Indirect Aggression:
Definition: to attack or injure indirectly through another person or object.
Items:
a. This boy tattles to the teacher about what other boys do.
b. Here is a boy who breaks things that belong to others.

In addition to the 13 items used to measure the various forms of aggression, the booklet contained seven filler items. The practical consideration of limited attention span to a written task in this age group necessitated the use of only a small sample of items for each aggression category.

A different form of question was used for two additional items in the booklet which made it possible to estimate the children's understanding of the "Guess Who" form of question. The children were asked to "List the 3 boys in your class who fight the most with other boys" and to "List the 3 boys in your class who say mean things, and threaten other boys the most." For the two different forms of item measuring Unprovoked Physical Aggression and Verbal Aggression, respectively, the average tetrachoric correlation coefficients for the five classroom groups were +.90 and +.88.

An overt aggression score was obtained for each *S* in each of the five aggression categories by counting the number of times his name was noted by both the boys and girls in the class. The Spearman rank-difference correlation coefficients (rho) between the boys' entries and the girls' entries for the five aggression measures ranged from +.87 to +.56; the entries were therefore combined to provide the five measures of aggression for each *S*.

The intercorrelations (the average rhos for the five classroom groups) among aggression variables are presented in Table 1. These correlations are largely positive in direction.

The aggression scores were derived from responses to verbally described classes of behavior and not from directly observed behavior. It is therefore possible that the children in making their judgments of each others' aggression were responding to the connotations of the verbal descriptions rather than on the basis of their actual observations of the behaviors described in the items. However, there were highly significant correlations between peer group ratings and teacher observations of the aggressive behaviors; biserial correlation coefficients between peer and teacher judgments for the five aggression variables ranged from +.80 to +.72 ($p < .01$ in all cases). There was therefore strong evidence that teachers and peers respond in

TABLE 1
INTERCORRELATIONS AMONG AGGRESSION MEASURES

	Provoked Physical Aggression	Outburst Aggression	Unprovoked Physical Aggression	Verbal Agression	Indirect Aggression
Provoked Physical Aggression					
Outburst Aggression	+.36				
Unprovoked Physical Aggression	+.33	+.63			
Verbal Aggression	+.27	+.73	+.73		
Indirect Aggression	−.21	+.36	+.43	+.57	

TABLE 2

CORRELATIONS BETWEEN AGGRESSION SCORES AND THE POPULARITY SCORE

	School A Grade 5 (N = 19)	School B Grade 5 (N = 10)	School B Grade 6 (N = 16)	School C Grade 5 (N = 16)	School C Grade 6 (N = 13)	Over-all Mean for All Classes (N = 74)
Provoked Physical Aggression vs. Popularity	+.47*	+.42	+.57*	+.10	−.02	+.31
Outburst Aggression vs. Popularity	−.43*	+.07	+.12	−.60**	−.23	−.21
Unprovoked Physical Aggression vs. Popularity	−.45*	−.22	+.05	−.62**	−.57*	−.36
Verbal Aggression vs. Popularity	−.48*	−.34	−.02	−.65**	−.75**	−.45
Indirect Aggression vs. Popularity	−.55**	−.56*	−.81**	−.68**	−.86**	−.69
Over-all Mean for All Aggression-Categories	−.29	−.13	−.02	−.49	−.49	−.28

* $P < .05$.
** $P < .01$.

common to the observed aggressive behaviors.

RESULTS

The Spearman rank-difference correlation coefficients (rho) between the five aggression variables and popularity are presented in Table 2 for each class separately.

The most striking feature of these results is the absolute consistency among all five classes in the progressively more negative correlations as the relationships between popularity and agression measures are considered in the following order: provoked physical aggression, outburst aggression, unprovoked physical aggression, verbal aggression, and indirect aggression. In each class, the correlation between provoked physical aggression is the most positive of the correlations for that class and the correlation between indirect aggression and popularity the most negative, with the correlations for outburst aggression, unprovoked physical aggression, and verbal aggression ordered in between as increasingly negative.

When all five aggression variables are considered simultaneously, the over-all relationship between aggression and popularity is negative for all classes. However, there are considerable differences among the classes in this respect. For example, the over-all mean correlation for all aggression categories for School B, Grade 6 is very nearly zero, the positive correlations for provoked physical aggression, outburst aggression, and unprovoked physical aggression balancing the negative correlations for verbal aggression and indirect aggression. In contrast, School C, Grade 6 manifests negative correlations for all aggression variables. Regardless of the size of the negative over-all relationship between aggression and popularity, however, the same sequence of progressively increasing negative correlations is maintained with absolute consistency for all classes.

DISCUSSION

This study presents evidence that, within the sample of *Ss* employed, there are clear differences in the lower-class peer group's reaction to different manifestations of aggressive behavior on the part of its members. Retaliation to an aggressive

attack is a relatively approved form of behavior and verbal aggression and indirect aggressive acts are strongly disapproved among a sample of preadolescent upper lower-class boys. However, these findings would probably not be duplicated for samples which differ in sex, age, and social class position. It would be of great value for the study of socialization to extend the knowledge of the parameters of peer group influence by utilizing a wide variety of samples differing, at least, for sex, age, socal class, religion, and geographic location.

The present sample occupies an upper lower-class position. The present finding that there is a generally negative relationship (although more negative for some forms of aggression than for others) between aggression and popularity among these upper lower-class boys is somewhat surprising in view of the statements and/or assumptions of previous studies (e.g., Davis, 1944; Mussen & Naylor, 1956). These findings qualify Davis' (1944) observation that "The lower classes not uncommonly teach their children and adolescents to strike out with fist or knife and to be certain to hit first.... The important consideration with regard to lower-class adolescents is that it is learned as an *approved and socially rewarded* form of behavior in their culture." (Davis, 1944, p. 209). Davis also remarks that even when lower-class parents discourage aggression, "...the power of the street culture in which the child and adolescent are trained overwhelms the parental verbal instruction. The rewards of gang prestige...seem to be on the side of the street culture." (Davis, 1944, p. 210). The results of the present study suggest not only that there are upper lower-class peer cultures that do not approve aggression in general but also that very precisely discriminated kinds of aggression are differentially rewarded or punished by upper lower-class children.

It is apparent, of course, that these results can not be generalized to peer behavior beyond the confines of the school. It is possible that there would be a greater resemblance between the results of this study and the observations of other investigators if measures of lower-class peer influence outside of the school had been obtained.

One additional point may be noted with regard to the ordering of the correlations between popularity and the various forms of aggression. For all five classes, outburst aggression is the second least disapproved form of aggression, while verbal aggression and indirect aggression are the most strongly disapproved by peers. This ordering reverses what appears experientially to be the response of adults to these forms of aggression in 10- to 13-year old boys. Teachers, for example, view outburst, uncontrolled, temper tantrum behavior as especially undesirable in a 10–13 year old boy,[3] while verbal and indirect forms of aggression are frequently permitted or encouraged by teachers as displaced, less destructive outlets for aggressive feelings. However, the reactions of peers to these aggressive responses is exactly the reverse. Outburst aggression is much less associated with a child's unpopularity with his peers than verbal and indirect aggression.

Summary

This study specifies the manner in which the acceptability of aggressive responses to a lower-class, preadolescent's peer group varies with the precise form of aggression manifested.

For five classroom groups of upper lower-class, preadolescent boys, five different forms of the overt display of aggression were measured. The correlations

[3] Beginning with Wickman's (1928) original study, a series of investigations of teachers' attitudes has consistently ranked "temper tantrums" among the most serious behavior problems in children.

between the various forms of aggression and popularity within the peer group indicates that among preadolescent boys Provoked Physical Aggression is relatively approved behavior, Outburst Aggression, Unprovoked Physical Aggression, and Verbal Aggression are progressively more disapproved, and Indirect Aggression is strongly disapproved.

REFERENCES

Appel, M. H. Aggressive behavior of nursery school children and adult procedures in dealing with such behavior. *J. exp. Educ.*, 1942, **11**, 185–199.

Cunningham, Ruth et al. Understanding group behavior of boys and girls. New York: Teachers College, Columbia Univer., Bureau of Publications, 1951.

Davis, A. Socialization and adolescent personality. *Adolescence, Forty-third Yearbook*, Part I. Nat. Soc. for Study of Educ. Chicago: Univer. Chicago Press, 1944.

Jersild, A. T., & Markey, F. V. Conflicts between preschool children. *Child Developm. Monogr.*, No. 21. New York: Teachers College, Columbia Univer., 1935.

Levin, H. The influence of classroom control on kindergarten children's fantasy aggression. *Elem. Sch. J.*, 1955, **55**, 462–466.

MacRae, D. A test of Piaget's theories of mental development. *J. abnorm. soc. Psychol.*, 1954, **49**, 14–18.

Murphy, Lois B. Social behavior and child personality. New York: Columbia Univer. Press, 1937.

Mussen, P. H., & Naylor, H. K. The relationships between overt and fantasy aggression. *J. abnorm. soc. Psychol.*, 1954, **49**, 235–240.

Piaget, J. The moral judgment of the child. London: Kegan Paul, 1932.

Sears, R. R., Whiting, J. W. M., Nowlis, V., & Sears, Pauline S. Some child-rearing antecedents of aggression and dependency in young children. *Genet. Psychol. Monogr.*, 1953, **47**, 135–234.

Tuddenham, R. D. Studies in reputation III. Correlates of popularity among elementary-school children. *J. educ. Psychol.*, 1951, **42**, 257–276.

Warner, W. L., Meeker, Marchia, & Eells, K. *Social class in America.* Chicago: Science Research Associates, 1949.

Whiting, J. W. M., & Child, I. L. Child training and personality: a cross-cultural study. New Haven: Yale Univer. Press, 1953.

Wickman, E. K. Children's behavior and teachers' attitudes. New York: The Commonwealth Fund, 1928.

Received July 8, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 1, 1959

SOME RELATIONS BETWEEN TEST SCORES AND ITEM STATISTICS

FRANCES SWINEFORD
Educational Testing Service

That there are definite relationships between score statistics for a group of examinees and item statistics based on data for the same group is well known. The most obvious one is that when the score is the number of correct responses the mean score is equivalent to the sum of the proportions of examinees who answer the items correctly. Gulliksen (1945) has proved that test reliability and score variance both increase (*a*) as the average inter-item correlation increases and (*b*) as the variance of the item difficulty distribution decreases. His presentation was concerned with theoretical relationships leading to a set of general theorems.

The present study was undertaken in the belief that simple empirical relationships which would serve to complement Gulliksen's theoretical formulas could be found. The results shed considerable light on the way in which item statistics and test statistics are likely to be related in any new test. Thus, reasonable estimates of item statistics can yield worthwhile predictions of score statistics. There will be presented regression equations and corresponding tables which can be used to estimate the score standard deviation, test reliability, and mean item-test correlation all within certain limits to be pointed out later.

The Variables

Two test-score statistics of particular interest are the standard deviation and the reliability. The score standard deviation, σ_t, is expressed in terms of score units, which vary from test to test. In order to eliminate this unit of measurement we have selected as our first variable the ratio, $(n - \text{chance})/\sigma_t$, where n is the number of items in the test and chance is the expected mean score if all answer sheets were marked at random. This ratio is a pure number. Although it does vary with test length, the variation is not great for reliable tests when the longest is no more than eight or ten times the length of the shortest. Such variation can be ignored for practical purposes.

The second variable is an estimate of test reliability. The reliability of the tests used in this study was computed by the Kuder-Richardson Formula [20] (1937, p. 158) for all tests for which Score = R and by the Dressel (1940, p. 309) adaptation of this formula for all tests for which Score = R − kW. In these formulas R represents the number of items which an examinee answers correctly, and W, the number answered incorrectly. As in the case of the standard deviation, it is necessary to make an adjustment in order to render coefficients for different tests comparable in some sense. Other things equal, long tests are more reliable than short tests. It was decided arbitrarily to adjust each observed reliability coefficient by means of the Spearman-Brown prophecy formula so that it would describe a 100-item test. This adjusted coefficient will be denoted rel._{100}.

The third variable, one of two item statistics to be used in the study, is a measure of the variability of item difficulty indices. The difficulty index used at Educational Testing Service, known as "delta," is a normal-curve deviate above which lies the area under the curve equal to the proportion of successful examinees. It is expressed in terms of a scale with mean of 13 and standard deviation of 4. Thus, for example, if 50 per cent of the examinees who attempt an item answer it correctly,

the corresponding delta is 13.0; if 60 per cent are correct, the delta is 12.0; and if 70 per cent are correct, the delta is 10.9. Delta is not linearly related to p, the proportion of correct responses to an item. It is used, however, because, for reasons not relevant to this study, the difficulties of ETS test items are recorded in terms of delta. The third variable, then, is the standard deviation of the n deltas, σ_Δ.

The fourth variable is the reciprocal of the square of the mean of the item-test correlations, $1/\bar{r}_{bis}^2$, which for simplicity will be denoted $1/\bar{r}^2$. C. T. Fan has shown that under certain conditions the relation between the Kuder-Richardson Formula [20] reliability and the function, $1/\bar{r}^2$, is essentially linear (Swineford, 1957). Item-test correlations are preferred to interitem correlations because they are more readily obtained. In a test of n items there are only n item-test correlations, whereas there are $n(n-1)$ interitem correlations. The inclusion of the item itself in the total score with which the item score is correlated produces in the correlation a spurious element, which is relatively higher for a short test than for a long one. Correcting each item-test correlation to eliminate this spurious element, however, is considered not worth the time and effort it would require.

The Populations

The populations, two in number, are multiple-choice tests. Tests scored by the formula, Score = R, comprise one population, and those scored by the formula Score = R − kW, comprise the other. If the test is composed of K-choice items, k is usually $1/(K-1)$. In some instances a "test" as used here is a separately scored and analyzed subtest from a larger examination. Each of the populations will be handled separately through a sample of tests administered by Educational Testing Service in such Programs as College Entrance Examination Board, Graduate Record Examinations, Law School Admission Test, National Board of Medical Examiners, and others. The data were taken from ETS' files. An effort was made to include in each sample a broad range of each variable. Greater success in this regard was achieved for Sample A, the "R" tests, than for Sample B, the "R − kW" tests. The tests vary in length from 15 to 260 items, with the majority within the range, 35 to 175.

The Data

The basic statistics are the means, standard deviations, and intercorrelations of the four variables. These are presented in Table 1. It should be noted that for Sample B Variable 1 becomes n/σ_t when every item has $1+1/k$ choices.

Variables 1, 2, and 4—those involving the score standard deviation, the test reliability and the item-test correlations—are interrelated to an extremely high degree. All the correlations for Sample B are lower than corresponding values for Sample A, as is consistent with the lower standard deviations. These differences are probably not attributable to the difference in the scoring procedure but principally to the fact that it was not possible to find in the files a wide representation of the four variables among tests scored R − kW.

The data of Table 1 will be used to compute regression equations for predicting the score standard deviation and the test reliability. It should be understood that the beta coefficients in the regression equations are not unique. For example, ETS-produced tests usually yield unimodal delta distributions centered near middle difficulty for appropriate examinee groups. The inclusion of tests with extremely leptokurtic or platykurtic or U-shaped delta distributions would undoubtedly have been reflected in the beta-coefficients.

Since it is sometimes convenient to be able to estimate the mean item-test correlation, given the score standard deviation and the test reliability, regression equations

TABLE 1
BASIC DATA FOR TWO TEST SAMPLES

Variable	Mean	Standard Deviation	Intercorrelations			
			1	2	3	4
Sample A: Score = R (74 Tests)						
1. $(n - ch)/\sigma_i$	6.2891	2.6847		−.9702	.6739	.9570
2. $rel._{100}$	0.8650	0.1025	−.9702		−.5859	−.9595
3. σ_Δ	2.2959	0.4391	.6739	−.5859		.6023
4. $1/r^2$	7.8250	5.4344	.9570	−.9595	.6023	
Sample B: Score = R − kW (54 Tests)						
1. $(n - ch)/\sigma_i$	5.1917	1.0171		−.8652	.2929	.9408
2. $rel._{100}$	0.9166	0.0379	−.8652		−.0157	−.9297
3. σ_Δ	2.0778	0.3089	.2929	−.0157		.0880
4. $1/r^2$	4.8000	1.8585	.9408	−.9297	.0880	

for this purpose will also be computed for both samples.

Prediction of Score Standard Deviation

The correlations show that the item-test correlations are far more closely related to the score standard deviation than is the variability of the item difficulty indices. The multiple regression equations are even more striking. In the equations which follow, z will be used to denote a variable in standard form, X will be used for the raw-score form, and a prime will denote a predicted value. The symbol, $R_{i(jk)}$, denotes the multiple correlation between X_i on the one hand and X_i predicted from X_j and X_k on the other.

The formulas in standard-score and in raw-score form for predicting Variable 1 = $(n - ch)/\sigma_i$ from Variable 3 = σ_Δ and Variable 4 = $1/r^2$ are as follows:

Sample A (Score = R)

$$z_1' = .1530\, z_3 + .8649\, z_4\,; \qquad [1]$$

$$X_1' = .9353\, X_3 + .4273\, X_4 + .7983; \qquad [2]$$

$R_{1(34)} = .9648$; $SE_{est.} = 0.706$.

Sample B (Score = R − kW)

$$z_1' = .2117\, z_3 + 9222\, z_4\,; \qquad [3]$$

$$X_1' = .6972\, X_3 + .5047\, X_4 + 1.3206; \qquad [4]$$

$R_{1(34)} = .9642$; $SE_{est.} = 0.270$.

The independent contributions of Variables 3 and 4 are proportional to the coefficients in Equations [1] and [3]. The contribution of σ_Δ is small, particularly in the case of Sample A. Table 2 gives values of $(n - ch)/\sigma_i$ for selected values of σ_Δ and $\bar{r}$ (the mean of the item-test biserial correlations). It should be noted that linear interpolation can be used within a column but not within a row of the table. From the tabled values it is readily seen that a very homogeneous test (high item-test correlations) produces such a large standard deviation that a score range of six standard deviations is very unlikely unless there is an unusually wide spread in item difficulty.

Prediction of Test Reliability (Kuder-Richardson)

The close relation between the Kuder-Richardson test reliability and the mean

TABLE 2

VALUES OF $(n - \text{chance})/\sigma_t$ FOR SELECTED VALUES OF σ_A AND $\bar{r}$

σ_A	$\bar{r}$						
	.20	.25	.30	.35	.40	.45	.50
Sample A (Score = R)							
3.5	14.8	10.9	8.8	7.6	6.7	6.2	5.8
3.0	14.3	10.4	8.4	7.1	6.3	5.7	5.3
2.5	13.8	10.0	7.9	6.6	5.8	5.2	4.8
2.0	13.4	9.5	7.4	6.2	5.3	4.8	4.4
1.5	12.9	9.0	6.9	5.7	4.9	4.3	3.9
1.0	12.4	8.6	6.5	5.2	4.4	3.8	3.4
0.5	11.9	8.1	6.0	4.8	3.9	3.4	3.0
Sample B (Score = R − kW)							
3.5	16.4	11.8	9.4	7.9	6.9	6.3	5.8
3.0	16.0	11.5	9.0	7.5	6.6	5.9	5.4
2.5	15.7	11.1	8.7	7.2	6.2	5.6	5.1
2.0	15.3	10.8	8.3	6.8	5.9	5.2	4.7
1.5	15.0	10.4	8.0	6.5	5.5	4.9	4.4
1.0	14.6	10.1	7.6	6.1	5.2	4.5	4.0
0.5	14.3	9.7	7.3	5.8	4.8	4.2	3.7

TABLE 3

VALUES OF rel._{100} FOR SELECTED VALUES OF $(n - \text{chance})/\sigma_t$ AND $\bar{r}$ AND FOR SELECTED VALUES OF $\bar{r}$ ONLY

$\frac{n-ch}{\sigma_t}$	$\bar{r}$						
	.20	.25	.30	.35	.40	.45	.50
Sample A (Score = R)							
14	.56						
12	.61	.67					
10		.72	.75				
8		.77	.80	.82			
7			.83	.85	.86	.87	
6			.85	.87	.88	.89	.90
5				.89	.91	.92	.92
4					.93	.94	.95
	.55	.72	.81	.86	.89	.92	.93
Sample B (Score = R − kW)							
	.53	.70	.80	.85	.89	.91	.93

item-test correlation is apparent from the data of Table 1. It happens that in the case of Sample A the use of Variable 1 as well as Variable 4 reduces the standard error of estimate by .007 from its value when Variable 4 is the sole predictor, whereas Variable 3 adds virtually nothing to the precision of the prediction. For Sample B, on the other hand, the use of Variables 1, 3, and 4 yields a standard error of estimate only .0003 less than that associated with Variable 4 alone. The following equations are suggested for practical use:

Sample A (Score = R)

$$z_2' = -.6178\, z_1 - .3683\, z_4\,; \qquad [5]$$

$$X_2' = -.02358\, X_1 - .006943\, X_4 + 1.0676; \qquad [6]$$

$R_{2(14)} = .9761$; $SE_{est.} = .022$.

TABLE 4

VALUES OF $\bar{r}$ FOR SELECTED VALUES OF $(n - \text{chance})/\sigma_t$ AND rel._{100}

$\frac{n-ch}{\sigma_t}$	rel._{100}							
	.60	.65	.70	.75	.80	.85	.90	.95
Sample A (Score = R)								
14	.21	.22						
12	.22	.23	.24					
10		.24	.25	.26				
8				.28	.30	.32		
7					.31	.34	.37	
6					.33	.35	.39	.44
5						.38	.42	.48
Sample B (Score = R − kW)								
14	.22	.23						
12	.23	.24	.25	.26				
10		.25	.26	.27	.29			
8				.30	.31	.33		
7					.33	.35	.38	
6						.38	.41	
5							.45	.51

Increase an estimate of the reliability is desired when only $\bar{r}$ is known, the following formula is offered for Sample A:

$$X_2' = -.01809\,X_4 + 1.0065; \quad [7]$$
$$R_{2(4)} = .9595; SE_{est.} = .029.$$

Sample B (Score = R − kW)

$$z_2' = -.9297\,z_4; \quad [8]$$
$$X_2' = -.01895\,X_4 + 1.0075; \quad [9]$$
$$R_{2(4)} = .9297; SE_{est.} = .014.$$

Values of rel.$_{100}$ for selected values of $(n - \text{ch})/\sigma_t$ and $\bar{r}$ are given for Sample A and values of rel.$_{100}$ for selected values of $\bar{r}$ are given for both samples in Table 3.

Estimation of the Mean Item-Test Correlation

When item-analysis data are not available, a reasonable estimate of the mean item-test correlation can be obtained from $(n - \text{chance})/\sigma_t$ and rel.$_{100}$ if the latter was computed by an internal procedure, such as one involving split-half correlations. The formulas for estimating Variable 4 = $1/\bar{r}^2$ from Variable 1 = $(n - \text{ch})/\sigma_t$ and Variable 2 = rel.$_{100}$ follow:

Sample A (Score = R)

$$z_4' = .448\,z_1 - .5280\,z_2; \quad [10]$$
$$X_4' = .9003\,X_1 - 28.0025\,X_2 + 26.38; \quad [11]$$
$$R_{4(12)} = .9655; SE_{est.} = 1.415.$$

Sample B (Score = R − kW)

$$z_4' = .5427\,z_1 - .4602\,z_2; \quad [12]$$
$$X_4' = .9916\,X_1 - 22.5720\,X_2 + 20.34; \quad [13]$$
$$R_{4(12)} = .9687; SE_{est.} = 0.461.$$

These formulas have been used to compute the entries in Table 4, which gives values of $\bar{r}$ for selected values of $(n - \text{ch})/\sigma_t$ and rel.$_{100}$.

Summary

Empirical methods have been employed to develop formulas for estimating the score standard deviation, the test reliability, and the mean item-test correlation from related statistics. The variables were kept as simple as possible, and only linear relationships have been considered, even though theory dictates the use of variables that are less readily obtained and relationships that are nonlinear. Nevertheless, the results show that the estimates can be made with a high degree of accuracy within the limits represented by the tests used. Without further evidence, it would probably be unwise to extrapolate beyond the values given in Tables 2, 3, and 4.

REFERENCES

DRESSEL, P. L. Some remarks on the Kuder-Richardson reliability coefficient. *Psychometrika*, 1940, **5**, 305-310.

GULLIKSEN, H. The relation of item difficulty and inter-item correlation to test variance and reliability. *Psychometrika*, 1945, **10**, 79-91.

KUDER, G. F., & RICHARDSON, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, **2**, 151-160.

SWINEFORD, F. Some relations between test scores and item statistics. *Res. Bull. No. RB-57-2*. Princeton, N. J.: Educational Testing Service, Feb., 1957. (Appendix)

Received July 16, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 1, 1959

THE COMPARATIVE PERSONALITY ADJUSTMENT OF SUPERIOR AND INFERIOR READERS

RALPH D. NORMAN AND MARVIN F. DALEY

University of New Mexico

In the voluminous literature on reading, there is still a paucity of reports on the comparative personality adjustment of good and poor readers. This situation exists despite the fact that such adjustment is reputedly associated with reading success. Almost 20 years ago, for example, Gates (1941) stated that emotional instability was found in about 75% of retarded readers. Yet, only recently Louttit (1957, p. 169) contended that "there is need for further research on the positive relationship between reading achievement and personality adjustment." Many previous studies have focused on the disabled reader himself without contrast with comparable controls; typically, these studies have used clinical or case-study methods. Such investigations are those by Vorhaus (1952), who examined reading disability cases with the Rorschach, and by others (Blanchard, 1928; Challman, 1939; Ephron, 1953; Kunst & Sylvester, 1943; Staiger, 1957; Stauffer, 1947; Wheeler, 1954) who utilized case-study approaches.

Many studies which have concentrated on comparison of good and poor readers generally have lacked certain precision. This lack has been due, in our opinion, to the fact that only fairly lately have sophisticated methods been introduced to deal with *patterns* or interaction effects in multivariable tests, and to the fact that researches have often dealt with subjective instruments, making it difficult to study differences in an exact manner.[1] While we do not minimize subjective approaches, more refined procedures hold out the possibility of greater exactitude in dealing with patterns. Indeed, a number of earlier comparative studies (Gann, 1945; Jackson, 1948; Spache, 1957) seem to support a hypothesis of personality *patterning* as related to reading disability.

It is the main purpose of the present study, therefore, to contrast the personality adjustment of superior and inferior readers as measured by a multivariable objective test with the aim of uncovering any discriminating psychometric patterns of adjustment. A second purpose is to provide a description of perceptions of themselves and their environment by both groups via an item analysis of this test.

METHOD

For the present study, all Anglo (non-Spanish) white males with an IQ between 84 and 116 on the Non-Language section of the California Test of Mental Maturity in the sixth grade in 14 middle class schools in Albuquerque, New Mexico, were scrutinized.[2] The restricted IQ range was chosen to cut out extremes in an attempt to control the factor of intelligence which correlates moderately with reading (Stephens, 1951). The Non-Language section was used since it is obviously less contaminated by reading ability than the Language section.

These pupils were examined for scores on the California Achievement Test to obtain two extremes in reading ability, i.e., those above or below one *SD* from their

[1] A recent exception is the study by Tabarlet (1958) who, however, while using an objective personality test, did not treat for psychometric patterns.

[2] These schools were chosen from the 50 in the city with consultation of guidance personnel who felt they were quite homogeneous with regard to socioeconomic status. We wish to thank Stanley Caplan, Associate Director of Guidance, for his assistance.

own grade level. The former were defined as superior readers ($N = 42$), all of whom had a reading grade placement score of 7.6 or better, and a mean of 8.1. The latter were defined as inferior readers ($N = 41$), all of whom had a grade placement of 4.8 or less, and a mean of 3.9. Thus, the average difference in reading skill between the groups was 4.2 grades. The mean Non-Language IQ of the superior group was 106.67; that of the inferior group was 98.85.

Examination of both groups on the basis of the criteria used (IQ and reading scores) produced the following results. An *AV* yielded an *F* between groups on reading grade placement of 1,267.24; $df = 1, 81$; $P < .01$. Another *AV* yielded an *F* (between groups on IQ's) of 18.10; $df = 1, 81$; $P < .01$. Correlations between IQ's and reading score for each group separately were not significantly different from zero.

Since we are interested in personality adjustment and since it was determined that there was a significant difference between the two groups in intelligence, the point may be raised concerning the possible effects of the latter on personality adjustment, i.e., that adjustment is associated with intelligence as well as with reading. If this were so, analysis based upon personality measurements would reflect results associated not with reading ability alone, but also with intelligence. However, two considerations regarding these three variables and their interrelationships should raise questions concerning the effects of intelligence. These are (*a*) that intelligence scores generally correlate poorly with personality scores, a point noted by Stephens (1951, p. 551) who states, "Typically, these investigations (on relationship between personality and intelligence) come out with extremely low correlations, ranging from −0.40 to 0.20 with an average close to zero," and (*b*) that in our study, from a *functional* standpoint, a difference of eight IQ points (in the middle range) between our groups is certainly insufficient to account for the tremendous disparity of four grades in reading achievement. Further, on this point, are studies (Bouise, 1955; Gann, 1945; Stewart, 1950) which matched *Ss* for intelligence and found adjustment differences between differentially able readers, using subjective instruments.

Both groups were administered the California Test of Personality (CTP), Elementary (Thorpe, Clark, & Tiegs, 1953), which has 12 subtests and in which language difficulty has been kept at or below fifth grade level.[3] (For profile purposes, the subtests have the advantages of high reliability and comparatively low intercorrelations.) Our general hypothesis is that the personality adjustment of superior readers will not differ significantly from that of the inferior; more particularly is the hypothesis that with a multivariable instrument (the CTP) the *patterns* of adjustment of both groups will not differ significantly. To test these hypotheses, *AV* technique applied to testing group psychometric patterns proposed by Block, Levine, & McNemar (1951) was employed. To perform item analyses for our second purpose, tables of Edgerton and Paterson (1926) for testing significance of differences between percentages were used.

Results

Table 1 gives the means and *SD*'s of the personality test scores of both groups.

[3] The reader who is concerned by this fact when our inferior group had a mean reading grade of 3.9 is reminded that *Ss* were instructed to ask for help in understanding any items. More importantly, the CTP authors claim that in the test standardization group, there were no significant differences between median scores of successive grade levels; if reading ability had a prominent effect on the CTP, one would expect grade changes in these average scores. Also, the test authors state that about 20% of the standardization samples were retarded one-half year or more.

The *AV* reveals an *F* ratio of 42.11 obtained for Groups; $df = 1, 81$; $P < .01$; and an insignificant *F* of less than one for Groups X Variables. Thus, our first hypothesis that the personality adjustment of the superior readers does not differ significantly from that of the inferior readers must be rejected; however, the null hypothesis of no difference between *patterns* of adjustment of the two groups cannot be rejected.

In line with the secondary purpose, it was found that 67 items of 144 on the test were significant at the 5% level or less in differentiating between the groups; answers in the indicated direction reveal a lower percentage for inferior readers in 64 of the 67 items.[4]

TABLE 1

MEANS AND STANDARD DEVIATIONS OF *T* SCORES ON CALIFORNIA TEST OF PERSONALITY VARIABLES FOR BOTH GROUPS

Variables	Inferior Readers		Superior Readers	
	Mean	*SD*	Mean	*SD*
Self-Reliance	46.90	6.77	52.38	7.27
Sense of Personal Worth	50.49	8.30	57.33	7.36
Sense of Personal Freedom	46.17	7.57	53.71	6.88
Feeling of Belonging	47.41	8.81	56.50	6.79
Withdrawing Tendencies (Freedom from)	51.32	7.84	59.38	6.96
Nervous Symptoms (Freedom from)	49.44	11.43	55.17	7.61
Social Standards	43.95	8.98	53.83	7.36
Social Skills	49.17	9.84	54.36	10.19
Anti-Social Tendencies (Freedom from)	44.80	10.21	52.14	9.44
Family Relations	45.12	8.47	52.52	6.83
School Relations	50.05	9.79	57.71	7.14
Community Relations	43.07	8.72	50.43	5.22

DISCUSSION

Whereas we have found no difference in *patterns* of adjustment between the two groups, there is a definite difference in *total* adjustment. Moreover, Table 1 reveals that there is a relatively constant difference of from five to 10 points between groups in the 12 test variables, the inferior readers demonstrating consistently poorer adjustment in all areas. We may then state that there is no difference in the *kind* of adjustment made by the two groups, considering kind to be reflected in patterns; but there is a strong difference in *degree* of adjustment. Thus, our results are only partially in agreement with Harris (1956, pp. 264–265) who states, "... research studies comparing the personalities of poor readers with good readers have failed to reveal any consistent group differences. This is probably due to the mistaken attempt to find a common personality type or problem in reading disability cases."

Although we recognize that some seven of 144 item comparisons should be significant by chance alone at the 5% level, there still remain 60 items reflecting real differences. It is interesting to note tendencies for these significant items to cluster together in a manner different from that based on the rationale of the 12 test variables, since they cut across them. In the discussion following, we have adopted a sort of "need-press" scheme of Murray (1938), selecting items most illustrative of these tendencies. A few items appear in more than one of these clinical clusters; six of the 67 do not appear, three of which (66, 67, 68) indicate inferior readers with significantly higher percentages of affirmative answers to questions on physical

[4] An 8-page table giving item analysis data has been deposited with the American Documentation Institute, Order Document No. 5853, from ADI Auxiliary Publication Project, Photoduplication Service, Library of Congress, Wash. 25, D. C. remitting $1.25 for 35-mm. microfilm or $1.25 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service.

symptoms, and three of which (94, 98, 140) are reversals.

Five clusters seem to describe the inferior reader's "presses" or perceptions of his external environment. *Poor Family Interaction* is suggested by one grouping which goes well beyond the test category of Family Relations because most of the items cited here fall outside that category. Two examples are:

27. May you usually bring your friends home when you want to? (66; 24)[6]
114. Do you like both of your parents about the same? (100; 73)

Ten other significant differences (Items 15, 21, 31, 34, 41, 46, 78, 81, 115, 119) show unstable home environment, family discord, conflict with and about parents or relatives, and authoritarian discipline.

A second cluster is *Rejection by Others* wherein items reflect feelings of being scorned, rebuffed or excluded from activities by peers and others. Two examples from this group are:

23. Do people often think that you cannot do things very well? (24; 61)
55. Are people often so unkind or unfair that it makes you feel bad? (19; 49)

Other items in this cluster are 14, 16, 43, 48, 53, 94, 121, 130, and 132.

A third grouping involves *Frustration-Aggression by Others*; here, maltreatment, quarreling, etc., instigated by others characterize the common element of the following examples and of other items (55, 60, 101, 103, 115, 129):

107. Do classmates often quarrel with you? (7; 37)
126. Does it seem to you that some of the teachers "have it in for" pupils? (12; 32)

Fourthly, a cluster of items centers in *Conflicts About Other-Dominance*, wherein exist perceptions of prohibitions, restraints, coming out second-best, etc. Examples below and Items 26, 31, 34, 78, 81, 85, and 88 are significant:

10. Do your parents or teachers usually need to tell you to do your work? (12; 49)
30. Are you prevented from doing most of the things you want to do? (9; 37)

A final group of "presses" depicts *Environmental Deprivation*, suggesting deficiencies of companionship, friendship, and interesting "things" and "places"; these items, besides the two below, are 32, 43, 58, 59, and 135:

45. Do you have just a few friends? (12; 37)
134. Do you think there are too few interesting places near your home? (24; 54)

Four categories refer more to internally expressed "needs" of the retarded reader. *Aggression Towards Others* is suggested in the following and other items (78, 82, 89, 99, 103); the common element appears to be overt or covert opposition to others, taking the form of resistance or more direct action:

74. Is it all right to disobey teachers if you think they are not fair to you? (0; 20)
101. Do people often act so mean that you have to be nasty to them? (14; 37)

Need for *Impulsivity in Action* wherein response is in accordance with impulse regardless of others, irresponsibility, and defiance of convention seems demonstrated by Items 1, 74, 77, 81, 82, 86, and 142 and these two examples:

73. Is it all right to cheat in a game when the umpire is not looking? (0; 15)
75. Should one return things to people who won't return things they borrow? (95; 66)

Rejection of Others appears to be the need common to the two examples below and to Items 77, 117, 125, and 143. This need is characterized by poor positive identifications, and abandonment or remaining indifferent to others:

[6] Numbers in parentheses after items refer to respective percentages of superior and inferior readers who answer the item affirmatively.

38. Are you proud of your school? (100; 90)
141. Do you dislike many of the people who live near your home? (12; 32)

Finally, there is a cluster which centers in *Inferiority Feelings*, encompassing eight items (12, 20, 24, 39, 52, 86, 88, 124). Examples:

24. Do most of your friends and classmates think you are bright? (71; 39)
88. Does it make you feel angry when you lose in games at parties? (2; 15)

The significant items showing "reversals" perhaps deserve a word of comment. These are:

94. Do the boys and girls seem to think you are nice to them? (51; 83)
98. Have unfair people often said that you made trouble for them? (50; 20)
140. Do you help children keep away from places where they might get sick? (29; 85)

We feel that, despite the retarded readers' strong feelings of rejecting and being rejected, there crops out in these items a longing to be thought well of by others, particularly peers. Item 140, in fact, has the second highest *CR* of all 67 significant items.

Our results are strikingly similar to those found by Tabarlet (1958) who compared 43 fifth graders, retarded two or more years in reading, with 29 average readers, all within average intelligence. Using a sociometric device, Tabarlet found that average readers chose more of their classmates and were chosen by the latter more often than were retarded readers; the differences were statistically significant. Our clusters of rejection of and by others among inferior readers correspond closely with this finding. Moreover, on an objective personality test other than the CTP, Tabarlet finds average readers scoring significantly higher than the retarded in categories of behavioral maturity, interpersonal skills, social participation, satisfying work and recreation, and adequate outlook and goals. Apparently, these outcomes are much like ours, even though our controls were superior rather than average readers.

Some comment is needed here about the possible effect of socioeconomic status on CTP results, especially since we selected our "middle class" schools on impression (see Footnote 2) and since Tabarlet reports that there appeared to be a tendency for more retarded readers to come from schools in lower socioeconomic areas. Some seven significant items (15, 32, 59, 117, 135, 142, and 134, cited above) might possibly support such a notion. On the other hand, there are eight items (33, 76, 79, 106, 110, 136, 139, 144), five of which deal directly with economic distress, which yield no significant differences between our two groups. It is indeed difficult to tell from the test itself whether the answer to items such as 134 springs from fact or the misperceptions of the retarded reader. We are prone to favor the latter viewpoint.

Design of the present study, of course, does not permit us to state whether personality problems are a cause or result of poor reading ability. Louttit (1957) cites numbers of studies whose authors make such a statement; our data simply confirm that the two factors are associated.

Summary

A group of 42 superior was compared with a group of 41 inferior readers; both groups were composed of sixth grade boys from middle class schools and had comparable mean IQ's. Mean reading achievement difference between groups was 4.2 grades. They were given the California Test of Personality in order to ascertain differential patterns of personality adjustment. Analysis of variance revealed no difference in pattern, but superior readers achieved significantly higher adjustment scores on all parts of the test. Inspection of 67 significant items suggested several clusters of "needs" and "presses" which differentiated between the two groups.

REFERENCES

BLANCHARD, PHYLLIS. Reading disabilities in relation to maladjustment. *Ment. Hyg.*, 1928, **12**, 772–788.

BLOCK, J., LEVINE, L. & MCNEMAR, Q. Testing for the existence of psychometric patterns. *J. abnorm. Soc. Psychol.*, 1951, **46**, 356–359.

BOUISE, L. M. Emotional and personality problems of a group of retarded readers. *Elem. Eng.*, 1955, **32**, 544–548.

CHALLMAN, C. C. Personality adjustment and remedial reading. *J. excep. Child.*, 1939, **6**, 7–12.

EDGERTON, H. A. & PATERSON, D. G. Table of standard errors and probable errors in percentages for varying numbers of cases. *J. appl. Psychol.*, 1926, **10**, 378–391.

EPHRON, BEULAH. *Emotional difficulties in reading*. New York: Julian Press, 1953.

GANN, EDITH. *Reading difficulty and personality organization*. New York: King's Crown Press, Columbia Univer., 1945.

GATES, A. I. The role of personality maladjustment in reading disability. *J. genet. Psychol.*, 1941, **59**, 77–83.

HARRIS, A. J. *How to increase reading ability*. New York: Longmans Green, 1956.

JACKSON, J. A survey of psychological, social, and environmental differences between advanced and retarded readers. *J. genet. Psychol.*, 1948, **65**, 113–131.

KUNST, MARY, & SYLVESTER, EMMY. Psychodynamic aspects of the reading problem. *Am. J. Orthopsychiat.*, 1943, **13**, 69–76.

LOUTTIT, C. M. *Clinical psychology of exceptional children*. New York: Harper, 1957.

MURRAY, H. A., ET AL. *Explorations in personality*. New York: Oxford Univer. Press, 1938.

SPACHE, G. D. Personality patterns of retarded readers. *J. educ. Res.*, 1957, **50**, 461–469.

STAIGER, R. C. Self-responsibility and reading. *Educ.*, 1957, **77**, 561–565.

STAUFFER, R. G. Clinical approach to personality and the disabled reader. *Educ.*, 1947, **67**, 427–435.

STEPHENS, J. M. *Educational psychology*. New York: Holt, 1951.

STEWART, R. S. Personality maladjustment and reading achievement. *Am. J. Orthopsychiat.*, 1950, **20**, 410–417.

TABARLET, B. E. A study of the mental health status of retarded readers. Unpublished doctoral dissertation, Louisiana State Univer., 1958.

THORPE, L. P., CLARK, W. W., & TIEGS, E. W. *California test of personality*. Los Angeles: California Test Bureau, 1953.

VORHAUS, P. G. Rorschach configurations associated with reading disability. *J. proj. Tech.*, 1952, **16**, 3–19.

WHEELER, L. R. Dealing with emotional problems in the classroom. *Educ.*, 1954, **74**, 566–571.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 1, 1959

CONCURRENT VALIDITY OF THE "WARM TEACHER SCALES"

M. STEPHEN SHELDON, JACK M. COALE, AND ROCKNE COPPLE
Colorado State College

In recent years a number of new psychological inventories have been published, and scales on existing inventories developed that purport to be helpful in selecting teachers and candidates to become teachers (Cook, Leeds, & Callis, 1952; Hathaway & McKinnley, 1943; Ryans, in press). In most cases empirical validity coefficients have been presented for these instruments. The criteria used for the validation were, in general, ratings made by supervisors, administrators, pupils, or trained observers. Such criteria may in themselves be questionable. Nevertheless the scales seem to be getting at some existent personality configuration that is rationally related to what a good teacher ought to be.

There are two indications that these teacher selection scales are each meauring about the same thing. First, the authors of several of the scales use almost synonymous adjectives in describing the attributes that are being measured, e.g., "warm," "friendly," "rapport building." Second, though the different scales use different types of items, (Likert, true-false, multiple choice), and the items themselves deal with different kinds of things, there is a relatively high correlation between the scales.[1]

If the assumptions are true that: Good teachers possess a particular personality structure and that many of the "warm teacher scales" are in some ways measuring this structure, then it should also hold true that a group of individuals attaining high scores on a number of these scales ought to look quite different on certain other psychological measures than a group attaining low scores on the same scales. It is the purpose of this paper to test that hypothesis and determine on which of several personality measures the high and low scores on the "warm teacher scales" differ.

[1] Marjorie Rapp, Selection and Counseling Service, School of Education, University of California at Los Angeles, computed correlations between the four MMPI teacher selection scales used in this study. The lowest of these correlations was of the order of .50. Further, Cook and Medley (1954), in comparing the Ho and Pv Scales, intentionally selected items which discriminated between individuals having high and low scores on the MTAI.

Procedures

The initial sample of *Ss* used in this study was comprised of 176 college freshmen enrolled in general psychology at Colorado State College. (The students were actually enrolled in five different sections of the course.) Each took the Minnesota Teacher Attitude Inventory (MTAI) (Cook et al., 1952) and responded to the items from the Minnesota Multiphasic Personality Inventory (MMPI) (Cook & Medley, 1954; Gowan & Gowan, 1955; Meehl & Hathaway, 1946) that are included on the K, Hostility (Ho), Pharisaic Virtue (Pv), and Teacher Prognosis (Tp), scales. Standard scores (T) were computed on each of the five scales and summed for each *S* in the sample. The 10 highest and 10 lowest cumulative scores were identified and all agreed to take part in further testing.

It was specifically hypothesized that the following measures would differentiate between the high and low scoring groups (henceforth identified as Group A and Group B, respectively): (*a*) the Study of Values (Allport, Vernon, & Lindzey, 1952), (*b*) six needs on the Edwards Per-

sonal Preference Schedule (EPPS) (Edwards, 1954), (*c*), the California *F* Scale (Ardono, Frankel-Brunswik, Levinson, & Sanford, 1950), (*d*) the Wechsler Adult Intelligence Scale (WAIS) (Wechsler, 1955), and (*e*) 10 cards from the Thematic Apperception Test (TAT) (Murray, 1943), scored in the manner suggested by Friedman (1957). These tests were administered to *Ss* in both Groups A and B.

For scoring the TAT a 60-item *Q* sort was devised (Stephenson, 1953). The 60-item sort included 10 items descriptive of each of six needs specified by Murray (1938), measured by the EPPS (Edwards, 1954), and rationally considered relevant in this study to differentiate between the A and B groups. Specifically measured were the needs for affiliation, nurturance, aggression, dominance, succorance, and abasement. The association of each item with the corresponding need was independently checked and agreed upon by a jury of four professional psychologists.

After each *S* had related his story in response to a TAT card the examiner asked that the *S* describe the hero of his story by throwing the *Q* sort. The forced distribution for this sort was a nine-point scale. This procedure was followed with *Ss* from Group A and B for each of 10 cards.

Since the *Q* sort was to be used by each *S* to describe 10 different TAT heroes, each item was a descriptive statement having no subject and a third person verb, e.g. "enjoys sympathy" (succorance), "is affectionate" (nurturance), "feels guilty when things go wrong" (abasement). Many of the items were direct quotations from the EPPS manual (Edwards, 1954).

The scores assigned by the *S* to the item associated with each need were summed for each sorting. This resulted in a score for each need for every TAT hero for all *Ss*. The total scores for each need were then averaged for individual *Ss*.

The significance of the difference between the means of Group A and Group B on each of the psychological measures was tested by computing *t* ratios.

Results

The results of the *t* test appear in Table 1.

The results in Table 1 show that the mean of Group A was significantly higher on the WAIS (more intelligent) and significantly lower on the California *F* Scale (less authoritarian).

No significant difference was found between Group A and Group B on any of the scales of the Study of Values.

As expected, on both the EPPS and the TAT, Group A demonstrated a significantly higher need for affiliation and a significantly lower need for succorance than did Group B. The significantly lower need for abasement demonstrated by Group A on the EPPS was not substantiated by the TAT measure of this same need.

Certain unexpected results occurred. On the TAT, Group A demonstrated a significantly lower need for dominance than did Group B and a significantly higher need for aggression than did Group B. On the EPPS these findings were reversed, Group A scoring a significantly higher need for dominance than did Group B and a significantly lower need for aggression.

Discussion

It appeared to the experimenters that there was some credence to be allowed their original hypothesis, i.e., groups of freshmen students at the institution where the experiment was conducted, when differentiated on the basis of the MTAI, and certain of the MMPI scales, were found to differ significantly on 11 of the 20 measures experimentally administered.

The direction of the significant differences found in this study will not be

TABLE 1

TABLE OF MEANS, MEAN DIFFERENCE AND *t* RATIOS FOR GROUP A AND GROUP B SCORES FOR EACH MEASURED VARIABLE

Test	Scale	Mean A	Mean B	Dif	*t*
WAIS		117	101	16	2.64*
Calif. F Scale		39.55	60.33	20.78	3.74**
Study of Values	Theoretical	39.00	40.50	1.50	.70
	Economic	39.78	39.00	.78	.48
	Aesthetic	37.67	40.00	2.33	1.23
	Social	37.00	36.00	1.00	.41
	Political	40.89	40.17	.72	.11
	Religious	45.89	43.17	2.72	1.14
TAT	Affiliation	56.78	53.17	3.61	2.31*
	Nurturance	56.22	55.33	.89	.18
	Aggression	51.33	46.17	5.16	3.11**
	Dominance	40.82	47.67	6.45	2.41*
	Succorance	46.11	53.67	7.56	3.01**
	Abasement	42.44	46.00	3.56	1.00
EPPS	Affiliation	18.56	15.83	2.73	2.16*
	Nurturance	14.00	16.17	2.17	1.44
	Aggression	8.33	11.50	3.17	3.21**
	Dominance	14.78	8.83	5.95	3.27**
	Succorance	9.56	14.33	4.77	2.74*
	Abasement	9.33	21.00	11.67	5.64**

* Significant at the 5% level of confidence.
** Significant at the 1% level of confidence.

surprising to those familiar with public education. It would be expected that friendly teachers and teacher candidates when compared to unfriendly ones would be more intelligent, less authoritarian, have a higher need for affiliation, and a lower need for aggression, succorance, and abasement.

Similar predictions could probably be made for some of the scales on the Study of Values. Certainly we would expect a group of warm teachers to have a higher social value than a group not so described. However, no significant differences were identified by this study for any of Study of Values Scales. It is altogether possible that, in spite of the rational expectancy, values so measured will not differentiate between friendly and unfriendly teachers. It is more probable that a major portion of the variance of the Study of Values Scales is accounted for by the values of the subcultures from which a sample of *Ss* is drawn. In the present study the sample seemed to be homogeneous in their subcultural backgrounds, i.e., rural midwesterners.

The reversals on the needs for dominance and aggression as measured by the EPPS and the TAT were not predicted and are difficult to account for. Several ex post facto explanations were explored in an attempt to cast light on these unique results. However, none was consistent with the other findings of the present study.

Several implications may be drawn from these data: (*a*) the so-called "warm teacher scales" seem to be measuring a

personality configuration which can be further analyzed by other psychological techniques, (*b*) a combination of such measures and scales may improve the prediction of those who select candidates for teacher training, (*c*) the positive results of this investigation warrant further study of the measurable personality characteristics of teachers designated as: "warm," "friendly," and "rapport building."

In regard to this latter point it is suggested that the above experimental conditions be repeated: (*a*) using a larger *N* and (*b*) with groups of student teachers and with teachers who have had some successful teaching experience.

Summary

A number of psychological scales, empirically developed, have been considered helpful in selecting teachers. The adjectives used for teachers who looked good on these scales are "warm," "friendly," "rapport building," etc. The present paper is concerned with how individuals, who score high and low on these scales, differ with respect to other psychological measures. The 10 persons scoring highest and the 10 lowest on the MTAI and four scales of the MMPI (K, Ho, Pv, and Tp) from a sample of 176 students at Colorado State College were given: (*a*) the Study of Values, (*b*) the EPPS, (*c*) the California F Scale, (*d*) the WAIS, and (*e*) the TAT.

The high and low groups were found to differ significantly in intelligence, authoritarianism, and certain manifest and latent needs.

REFERENCES

ADORNO, T. W., FRANKEL-BRUNSWICK, ELSE, LEVINSON, D. J. & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ALLPORT, G. W., VERNON, P. E., & LINDZEY, G. *Study of values*. New York: Psychol. Corp., 1952.

COOK, W. W., LEEDS, C. H., & CALLIS, R. *Minnesota teacher attitude inventory*. New York: Psychol. Corp., 1952.

COOK, W. W., & MEDLEY, D. M. Proposed hostility and pharisaic-value scales for the MMPI. *J. appl. Psychol.*, 1954, **38**, 414–418.

EDWARDS, A. L. *Edwards Personal Preference Schedule*. New York: Psychol. Corp., 1954.

FRIEDMAN, I. Objectifying the subjective—A methodological approach to the TAT. *J. proj. Tech.* 1957, **23**, 243–247.

GOWAN, J. C., & GOWAN, M. S. A teacher prognosis scale for the MMPI. *J. educ. Res.* 1955, **49**, 1–12.

HATHAWAY, S. R., & MCKINNLEY, J. C. *Minnesota multiphasic personality inventory*. New York: Psychol. Corp., 1943.

MEEHL, P. E., & HATHAWAY, S. R. The K-factor as a suppressor variable in the MMPI. *J. appl. Psychol.* 1946, **30**, 525–564.

MURRAY, H. A. ET AL. *Explorations in personality*. New York, London: Oxford Univ. Press, 1938.

MURRAY, H. A. ET AL. *Thematic apperception test*. Cambridge: Harvard Univer. Printing Office, 1943.

RYANS, D. G. Report on the teacher characteristics study. Wash. D. C.: Amer. Council on Educ., in press.

STEPHENSON, W. *The study of behavior*. Chicago: Univer. Chicago Press, 1953.

WECHSLER, D. *Wechsler Adult Intelligence Test*. New York: Psychol. Corp., 1955.

Received August 22, 1958.

JOURNAL OF EDUCATIONAL PSYCHOLOGY
Vol. 50, No. 1, 1959

AN EVALUATION OF THE EFFECTIVENESS OF A FRESHMAN MATHEMATICS COURSE[1]

J. STANLEY AHMANN AND MARVIN D. GLOCK
Cornell University

The purpose of this study is to evaluate the effectiveness of a one-semester experimental course in applied mathematics designed for college freshmen. The *Ss* used were freshmen enrolled in the College of Agriculture at Cornell University during the years 1952 through 1955. The course was evaluated in terms of their subsequent achievement as represented by grade-point averages, final marks in courses involving mathematics, scores on a mathematics test, and tendency to remain enrolled at Cornell University.

METHOD

Specifically, the effectiveness of the experimental course in applied mathematics was evaluated by comparing an experimental and control group in terms of (*a*) subsequent achievement at Cornell University (groups not stratified), (*b*) subsequent achievement at Cornell University (each group stratified into two subgroups in terms of high school grade-point average), (*c*) gains in knowledge of mathematics as measured by the Cornell Mathematics Test, and (*d*) tendency to remain enrolled at Cornell University. Although both four-year and two-year students were involved in the evaluation, only the four-year students are included in this report.

[1] This research project has been a cooperative one. In addition to the authors, contributions have been made by many other members of the staff of the Cornell University Testing and Service Bureau between 1947 and 1958. Although not a member of the Bureau staff, H. A. Geiselmann of Cornell University was a very active participant in the middle and late stages of the project and wrote a doctoral thesis concerning it. Some of his findings are included in the following tables.

Cornell Mathematics Test

The 1952 edition of the Cornell Mathematics Test, hereafter referred to as the CMT, is essentially a power test which measures achievement in arithmetic and basic algebra. It is composed of 55 multiple-choice test items and requires two and one-half hours of testing time.

The degree to which this test was valid was established in several different ways. First, an attempt was made to guarantee a high degree of content validity by following a carefully designed plan of test construction. The initial step was to construct sample test items covering most of the common arithmetic and algebraic operations. These were then sent in the form of a questionnaire to 101 College of Agriculture faculty members and to instructors in the physical sciences who taught many agriculture students. They edited the items and added others. Each item was then rated in terms of the importance, insofar as their courses were concerned, of the mathematical operations needed to solve it. These results were used to construct the 1951 edition of the test. Administrations of this edition provided additional data about item difficulty and item discriminating power; this information was used to develop the 1952 edition. The correlation between this test and the Cooperative Mathematics Pre-Test for College Students was 0.71 (Geiselmann, 1955, pp. 67–68).

Additional unpublished studies to establish the degree of validity of the CMT concerned concurrent validity. Students who had completed advanced algebra in high school significantly surpassed those who had completed only intermediate algebra in terms of this test. Also, those who had

completed intermediate algebra significantly surpassed those who had completed only elementary algebra. In addition, the means of the mathematics test scores of students who thought their high school mathematics preparation was good, surpassed the means of those who thought the opposite. Finally, seniors significantly surpassed freshmen in terms of the CMT after individual differences in scholastic aptitude and prior academic achievement had been controlled. The last comparison included an analysis of covariance in which the Ohio State Psychological Examination scores and the Cooperative General Science Test scores were used as controls.

Several reliability estimates were made. The most recent computation using the Kuder-Richardson Formula #20 yielded an r value of 0.89 (Geiselmann, 1955, p. 66).

Experimental Course in Applied Mathematics

The content of the applied mathematics course was selected primarily on the basis of the item-by-item analysis of the student responses to the CMT, suggestions from members of the College of Agriculture faculty, and analyses of the textbooks used by the students in agriculture and physical science courses. The principal topics covered were computation with a slide rule, fundamentals of algebra, equations, significant figures and approximation, ratio and proportion, trigonometry of the right triangle, graphs, and logarithms (Geiselmann, 1956). All of these topics are typically a part of junior high school arithmetic and elementary algebra programs.

The course was taught three hours per week for 15 weeks during the fall term. Each term there were six sections of about 25 students per section. The same instructor taught all sections all terms.

Selection of Subjects

The top 20% of each incoming freshman class in terms of mathematics skill as measured by the CMT was excluded from the study. Because of class scheduling problems, this group had to be identified before this test could be administered. In anticipation of this, earlier samples of freshmen were tested, and a regression equation was computed in which the number of high school mathematics courses, the high school grade-point average in mathematics, and the high school grade-point average, not including mathematics, were used to predict CMT-scores. To obtain the samples used in this present study, predictions were made for each entering freshman and he became a part of the sample or was rejected according to the size of his predicted CMT-score. Also excluded from the study, irrespective of the size of their CMT-scores, were all women students, transfer students, foreign students, and part-time students. Half of the freshmen in the eligible group were enrolled in the applied mathematics course and served as the experimental group; every second student was placed in the control group which did not receive any formal instruction in mathematics.

Results

Subsequent Achievement (Groups not Stratified)

The experimental group was compared with the control group in terms of six cumulative grade-point averages and final marks in eight introductory courses involving mathematics. In each instance an analysis of covariance (Wert, Neidt, & Ahmann, 1954, pp. 343–363) was computed. Two control variables were used, viz., the Ohio State University Psychological Examination and the CMT. The results for a three-year period are summarized in Table 1.

In only two instances were the differences between the means significant. Both involved final marks in chemistry courses. In one case the difference between the means favored the experimental group; in the second case it favored the control group. In summary, the influence of the experi-

TABLE 1

COMPARISON OF EXPERIMENTAL AND CONTROL GROUPS BY MEANS OF ANALYSES OF COVARIANCE IN TERMS OF SUBSEQUENT ACHIEVEMENT

Criterion	*N*		Mean		*F* value
	Exp.	Con.	Exp.	Con.	
Grade-point averages at end of:					
One Semester	271	255	72.0	71.7	2.47
Two Semesters	246	221	72.5	73.4	1.39
Three Semesters	112	105	74.6	75.4	1.13
Four Semesters	99	90	74.9	75.8	1.15
Five Semesters	51	36	75.5	76.3	0.64
Six Semesters	50	34	75.9	76.9	0.99
Final Marks in:					
Chemistry 105	119	121	64.0	61.2	3.98*
Chemistry 106	83	74	67.2	67.7	0.14
Chemistry 101	152	133	68.1	70.9	2.91
Chemistry 102	116	111	68.3	71.7	6.12*
Agr. Engr. 1	52	89	76.2	74.7	2.72
Physics 103	55	36	67.9	66.9	0.44
Physics 104	43	27	68.3	69.2	0.04
Mathematics 161	9	18	65.0	68.1	0.01

* Significant at .05 level.

mental course could not be detected to any important degree in subsequent Cornell achievement when the two groups were not stratified.

Subsequent Achievement (Groups Stratified)

Both the experimental and control groups were stratified into two subgroups in terms of high school grade-point average. The upper group in each case had averages of 85 or above; the lower group had averages of below 85. The experimental subgroups were compared to their corresponding control subgroups in terms of two Cornell grade-point averages and final marks in seven introductory courses in chemistry, engineering, and physics. In each instance an analysis of covariance was computed using the same two control variables mentioned earlier. The results for a three-year period are summarized in Table 2.

Again only two significant differences were found, one favoring the experimental group and one the control group. In the first case the criterion was final marks in chemistry and in the second case final marks in agricultural engineering. As in the case of Table 1 the results shown in Table 2 reveal little or no improvement in subsequent achievement which can be attributed to the experimental course in applied mathematics.

Gain in Mathematical Knowledge

The CMT was administered to both experimental and control groups as a pretest in late September and again as a final test in early May, over seven months later. Each student's gain was computed by finding the difference between his final test score and his pretest score. By means of an analysis of variance it was found that the mean gain of the experimental group (5.5) significantly surpassed the mean gain of the control group (2.2).

Tendency to Remain in College

The experimental and control groups were compared in terms of the tendency of

TABLE 2

COMPARISON OF STRATIFIED SUBGROUPS OF EXPERIMENTAL AND CONTROL GROUPS BY MEANS OF ANALYSES OF COVARIANCE IN TERMS OF SUBSEQUENT ACHIEVEMENT

Criterion	Subgroup	*N*		Mean		*F* value
		Exp.	Con.	Exp.	Con.	
Grade-point averages at end of:						
One Semester	Upper	129	143	73.8	73.4	0.89
	Lower	142	103	70.4	69.7	1.01
Two Semesters	Upper	121	130	74.2	74.6	0.12
	Lower	125	89	70.9	71.9	0.77
Final Marks in:						
Chemistry 105	Upper	71	71	65.8	64.1	1.09
	Lower	40	40	60.1	57.0	1.88
Chemistry 106	Upper	52	48	68.3	70.3	1.08
	Lower	26	24	65.4	62.6	1.42
Chemistry 101	Upper	60	66	69.7	73.2	1.95
	Lower	92	63	67.0	69.1	1.22
Chemistry 102	Upper	49	62	70.7	72.1	0.16
	Lower	67	49	66.3	71.1	7.88*
Agr. Engr. 1	Upper	42	74	80.6	77.4	4.49*
	Lower	48	44	73.8	72.8	0.54
Physics 103	Upper	29	25	69.7	69.6	1.16
	Lower	26	11	65.9	60.5	2.81
Physics 104	Upper	23	20	69.7	70.6	1.44
	Lower	20	7	66.6	64.9	0.17

* Significant at .05 level.

their members to remain enrolled in Cornell University. These comparisons were made at the end of the semester in which the applied mathematics course was taught, at the end of the first semester following, and again at the end of the second semester following. Chi-square analyses were computed for each semester individually and for the semesters cumulatively. A dropout was defined as a student who left the University with a cumulative grade-point average of less than 70.

The results for a three-year period are summarized in Table 3. The experimental group contained 278 students, the control group, 269. In two of the comparisons significant chi-square values were found. These and the nonsignificant values show that the attrition rate was initially much greater for the control group than it was for the experimental group. Later it was greater for the experimental group. By the end of the second semester following the one in which the mathematics course was taught, the cumulative attrition for the two groups was almost the same, about 25%. It seems that the mathematics course was instrumental in delaying the semester in which a student might leave the University but did not reduce the over-all tendency to drop out.

Student and Faculty Opinion

Students in the experimental group were asked their opinion of the mathematics course following the completion of it. Their faculty advisers asked them whether such a course should be offered. Sixty-six per cent responded affirmatively. About the same percentage of the faculty advisers responded in the same manner. In addition, a questionnaire administered by the instruc-

tor at the end of the mathematics course showed that 51% of the experimental group thought that the course would be worthwhile for all agriculture freshmen, 91% thought that it would be worthwhile for those poorly prepared in mathematics, and 78% thought that they were at least fairly well prepared in mathematics because of their work in the course. In the opinion of the majority of the students in the experimental group and their advisers, the mathematics course was a success.

Discussion and Conclusions

Certainly the foregoing results do not indicate that the experimental course in applied mathematics made a pronounced impact on the subsequent academic success of agriculture students. It is possible that a part of the influence of the experimental mathematics course was nullified by the manner in which the courses in agriculture and physical sciences were taught. Because in past years the students were known to be deficient in mathematics, some of the instructors supplemented their courses with small amounts of mathematics training considered to be necessary. No doubt this practice continued during the evaluation of the mathematics course. If so, both experimental and control group students received this instruction. The mathematics course is being continued and attempts to evaluate its effectiveness with respect to specific types of students (e.g., overachievers and underachievers) are being made.

Summary

The effectiveness of an experimental course in applied mathematics for agriculture freshmen (except those whose mathematical skill is superior) was evaluated by comparing an experimental and a control group in terms of subsequent grade-point averages, final marks in courses involving mathematics, gains in knowledge of mathematics, and tendency to remain enrolled at Cornell. In addition, the opinions of the students enrolled in the course and their faculty advisers concerning the course were obtained.

The statistical evidence does not indicate that the experimental course influenced subsequent achievement or ultimate attrition rate to any important degree. However the student and faculty opinions concerning the worth of the course were favorable.

TABLE 3

Comparison of Experimental and Control Groups by Means of Chi-square Analyses in Terms of Tendency to Dropout

Semester	Group	No. of drop-outs	Percentage of drop-outs	Chi-square value
Individual:				
During semester of the course	Exp.	7	2.5	
	Con.	13	4.8	2.1
During first semester following	Exp.	14	5.0	
	Con.	26	9.7	4.3*
During second semester following	Exp.	50	18.0	
	Con.	34	12.6	3.0
Cumulative:				
End of first semester after course	Exp.	21	7.6	
	Con.	39	14.5	3.9*
End of second semester after course	Exp.	71	25.5	
	Con.	73	27.1	0.2

*Significant at the 5% level.

References

Geiselmann, H. A. The effectiveness of a mathematics review course for freshmen in the College of Agriculture at Cornell University. Unpublished doctoral dissertation, Cornell Univer., 1955.

Geiselmann, H. A. Mathematical deficiencies of college freshmen. *Math. Teach.*, 1956, **49**, 22–29.

Wert, J. E., Neidt, C. O., & Ahmann, J. S. *Statistical methods in educational and psychological research*. New York: Appleton-Century-Crofts, 1954.

Received October 4, 1958.

to show how counseling can be one of the most significant influences on the personal development of the growing individual—

COUNSELING FOR PERSONAL ADJUSTMENT IN SCHOOLS AND COLLEGES

Fred McKinney

584 pages 1958 $6.00

to help the teacher help the child—

MENTAL HYGIENE IN ELEMENTARY EDUCATION

Dorothy Rogers

497 pages 1957 $5.50

to cover those recent advances which lead to a better understanding of child behavior and facilitate growth and learning—

READINGS IN EDUCATIONAL PSYCHOLOGY

edited by Jerome M. Seidman

402 pages 1955 Paper covers $3.25

HOUGHTON MIFFLIN COMPANY

Boston 7 Massachusetts | New York 16 New York | Atlanta 5 Georgia | Geneva Illinois | Dallas 1 Texas | Palo Alto California

Recently Published

FUNDAMENTALS OF MEASUREMENT

Techniques and Practices

by N. M. DOWNIE, *Purdue University*

An elementary textbook presenting the basic principles of educational and psychological measurement with emphasis on those which are fundamental to *all* areas of learning. Concise and practical, the text is ideally suited to the one-semester course offered in most colleges and schools of education.

Part I is an introduction to statistics, scores, theory, and the general use of tests. Part II offers more detailed treatment of achievement tests. Part III includes chapters on the use and interpretation of measurement and appraisal techniques. The book contains numerous charts and tables, a bibliography, a complete index, and a glossary of terms.

1958 424 pages; illustrated $6.00

OXFORD UNIVERSITY PRESS

417 Fifth Avenue, New York 16, New York

OPUS #1

An event of importance for

NEW AUTHORS ONLY

OPUS #1 is the title of a new Philosophical Library 1959 book series which will publish the deserving manuscripts of new authors only.

OPUS #1 has two basic requirements: The author must have something to say and know how to say it.

OPUS #1 will be dedicated to the humanities and will include works in the fields of Art, Literature, Philosophy, Religion, Sociology, History, Psychology and Psychiatry.

OPUS #1 books will be printed on special, fine antique paper, and artistically cloth-bound for library use.

Only complete book-length manuscripts accepted.

New Publications

Schizophrenia Manfred Sakel	5.00
Epilepsy Manfred Sakel	5.00
The Analysis of Dreams Medard Boss	6.00
Group Psychoanalysis B. Bohdan Wassell	3.75
The Neuroses and their Treatment Edward Podolsky	10.00
Psychotherapy and Society Wladimir Eliasberg	6.00
The New Chemotherapy in Mental Illness H. L. Gordon	12.00
Principles of Self-Damage Edmund Bergler	6.00
The New Psychiatry Nathan Masor	3.75
Experimental Psychology Ivan Pavlov	7.50

Inquire for full details

Philosophical Library, Publishers

Dept. OP-110, *15 East 40 St., N.Y. 16, N.Y.*

SCHOOLS OF PSYCHOLOGY

A PLAY
FOR CLASSES IN GENERAL PSYCHOLOGY AND EDUCATIONAL PSYCHOLOGY

By M. M. Kostick, Ed. D.

Head, Psychology Department
State Teachers College, Boston

"Although differing points of view may be stimulating and challenging to the professional psychologist, all too often the beginning student becomes confused and discouraged as he is confronted with the thinking of various schools of psychology. Dr. M. M. Kostick's little play, *Schools of Psychology*, goes far in helping the student see the bigger aspects and problems of psychology. The play presents what seems to be conflicting material in a manner which will maintain interest during the class period. The table at the close of the play summarizes the material in a form conducive to later serious study."

—T. L. Engle, Associate Professor of Psychology, *Indiana University*, and author of *Psychology* (3rd ed.), published by World Book Company

Price 75¢

Published by
Applied Psychology Associates
1658 Beacon St.
Brookline 46, Mass.

Copyright 1959 M. M. Kostick

Publication Manual

of the American Psychological Association

1957 Revision

A revision of the 1952 Manual, detailed instructions are given for the preparation of scientific articles. Organization and presentation of tabular material, figures and graphs, and reference lists are included. Scientists writing for publication will find the Publication Manual an indispensable guide.

Price, $1.00

Discounts for quantity orders over fifty copies

Order from

AMERICAN PSYCHOLOGICAL ASSOCIATION

Publications Office
1333 Sixteenth Street, N. W.
Washington 6, D. C.

A NEW AND IMPORTANT CONTRIBUTION IN EDUCATIONAL PSYCHOLOGY BY FREDERICK J. McDONALD, STANFORD UNIVERSITY

"The author of this new book has something to communicate and his ideas—and the ways in which they 'make sense'—are going to be appreciated. The book is systematic, better articulated, and more maturely presented than most other books in the field. McDonald's *Educational Psychology* will stand as a quality product."

The above is from one of the first reviews of this new text. The author's frank emphasis on a critical, conceptual, hypothesis-making approach to the psychological foundations of learning will achieve a major goal: aiding new teachers in becoming thoughtful and critical generators and testers of hypotheses.

This new book enables the student of teaching and psychology to find new ways of looking at human behavior—particularly the behavior of the child in the classroom—by providing a basic conceptual framework into which the student can integrate new psychological information, generalizations, hypotheses, and theories.

Frederick J. McDonald's *Educational Psychology*—in centering upon educational processes as they affect the total personality—is already setting trends in the field.

You are invited to write Box 31 for a first-press review copy.

A NEW BOOK FROM WADSWORTH PUBLISHING COMPANY • 431 CLAY STREET • SAN FRANCISCO

Good News ~

for the hundreds of teachers who have found it "the best, the most interesting, the most useful text on the subject"—

the new 5th edition of

Psychology of Adolescence

by

LUELLA COLE

revised in collaboration with IRMA N. HALL

will be ready in April

You'll like the new organization, bringing together in a final section the specific applications to teaching of concepts discussed earlier; the interesting new chapter on the goals of today's adolescents; the reports of the latest data and ideas from all related fields in the new edition of the text always considered outstanding for its well balanced presentation of all aspects of adolescent growth: physical, intellectual, emotional, social, and moral; normal and deviant.

from the ® *publisher*

of these interesting recent books

THE CHILD: A Book of Readings

by Jerome Seidman

65 articles by psychologists, doctors, educators, biologists and others present the major ideas and significant data on child growth.

ASSESSMENT OF HUMAN MOTIVES

edited with introduction by Gardner Lindzey

8 noted psychologists present contemporary thought on and investigative techniques for this central psychological problem.

STATISTICAL ANALYSIS

Revised Edition

by Allen L. Edwards

"A masterpiece of clarity," this basic text makes thoroughly understandable the principles and techniques of statistics needed by educators, psychologists and others. Accompanying *Workbook* available.

FIRO: A 3-Dimensional Theory of Interpersonal Behavior

by William C. Schutz

An interesting new theory of group behavior; its supporting data and applications.

For copies, write to

RINEHART

& Company, Inc. • 232 Madison Ave., New York 16